

**MODELO PARA LA RECUPERACION DE LA INFORMACION EN REDES
SOCIALES ONLINE BASADO EN EL PERFIL DE USUARIO Y
ONTOLOGIAS**



**Iván Darío Cerón Moreno
Jhon Alberto Astaiza Perafán**

Director: Magister. Miguel Ángel Niño Zambrano

Asesor: PhD(C). Carlos Cobos Lozada

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Sistemas
Grupo I+D en Tecnologías de la Información
Popayán, Enero de 2012**

**MODELO PARA LA RECUPERACION DE LA INFORMACION EN REDES
SOCIALES ONLINE BASADO EN EL PERFIL DE USUARIO Y
ONTOLOGIAS**



**Iván Darío Cerón Moreno
Jhon Alberto Astaiza Perafán**

Monografía presentada para optar al título de Ingeniero de Sistemas

Director: Magister. Miguel Ángel Niño Zambrano

Asesor: PhD(C). Carlos Cobos Lozada

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Sistemas
Grupo I+D en Tecnologías de la Información
Popayán, Enero de 2012**

NOTA DE ACEPTACIÓN

PRESIDENTE DEL JURADO

JURADO

AGRADECIMIENTOS

Quiero agradecer a Dios, a mi madre Luz Mila Moreno por su amor y constante lucha, a mi padre Teodulio Cerón por siempre estar presente y por ser mi ejemplo a seguir, a mis hermanos Kelly Cerón y Eduardo Cerón por su apoyo y confianza y a mis tíos y primos, por estar siempre pendientes de mi.

Agradezco también a mis sobrinos, a Marcela Velasco y mi hija Sara Cerón Velasco, quienes desde su llegada se convirtieron en una motivación más para el desarrollo de este proyecto.

Gracias a mi compañero de tesis Jhon Alberto Astaiza, por su disposición y colaboración permanente para este trabajo de grado.

Doy gracias al Ingeniero Miguel Ángel Niño por la dirección de este proyecto, sus asesorías y colaboración en la realización del mismo. Igualmente al ingeniero Carlos Cobos por su disposición y ayuda en los momentos necesarios.

A mis compañeros de universidad y sus familias, por su apoyo constante y por los momentos compartidos a lo largo de la carrera y en esta investigación.

Y finalmente muchas gracias a todas las personas que de una u otra forma, participaron en la realización de este proyecto.

Iván Darío Cerón Moreno

Quiero agradecer inicialmente a mis hermanos y hermanas por su amor y apoyo incondicional a lo largo de todos estos años que he compartido a su lado, en especial a mis hermanas Elizabeth Astaiza Perafán, Disa Amanda Astaiza, Nieves del Mar Astaiza Perafán y Mirta Jazmín Astaiza por ser los pilares fundamentales de mi familia y demostrar que es posible triunfar en este mundo pese a todos los inconvenientes y retos que nos presenta la vida. Igualmente agradezco a mi hermano Sergio Andrés Astaiza Perafán, futuro profesional del Diseño Gráfico, por sus valiosos aportes en el diseño de la interfaz de la aplicación Web, también a mis sobrinos que son esas personitas que te sacan una sonrisa del alma cuando las cosas parecen no tener rumbo.

Agradezco a mi padre Víctor Alberto Astaiza, por su amor, confianza, apoyo y por su enorme esfuerzo para lograr que cada uno de mis hermanos, se formen como personas y como profesionales. A mi madre Rudecinda Perafán, por ese amor, sabiduría y carisma infinitos que me brindo durante tantos años, los cuales me formaron como persona y como futuro profesional, y aunque hace ya dos años abandono este mundo terrenal, toda esa energía pura y hermosa que irradió a lo largo de su vida me acompaña y lo seguirá haciendo hasta el final de mis días.

Agradezco a mi compañero de tesis, Iván Darío Cerón Moreno, por su colaboración y disposición para este trabajo de grado.

Gracias al Ingeniero Miguel Ángel Niño por la dirección de este proyecto, sus asesorías y colaboración en la realización del mismo. Igualmente al ingeniero Carlos Cobos por su disposición y ayuda en los momentos necesarios.

Jhon Alberto Astaiza Perafán

TABLA DE CONTENIDO

1	INTRODUCCION	1
1.1	CONTEXTO GENERAL	1
1.2	DECLARACION DEL PROBLEMA	2
1.3	ESCENARIO DE MOTIVACION	3
1.4	CONTRIBUCIONES	3
2	MARCO TEORICO	5
2.1	ONTOLOGIAS	5
2.1.1	Componentes de una Ontología	5
2.1.2	Tipos de Ontologías	7
2.1.3	Metodologías para la Construcción de Ontologías	7
2.1.4	Herramientas para el Desarrollo de Ontologías	9
2.2	PERFILES DE USUARIO	10
2.2.1	Representación del Perfil de Usuario	10
2.2.2	Perfiles de usuarios basados en Ontologías	13
2.3	REDES SOCIALES	15
2.3.1	Tipos de Redes Sociales	15
2.3.2	Características de las Redes Sociales	15
2.3.3	Uso y motivaciones en las RSO	19
2.4	RECUPERACION DE LA INFORMACION	20
2.4.1	Modelos de Recuperación de información	20
2.4.2	Técnicas para la Recuperación de la Información.	20
2.4.3	Evaluación en la Recuperación de la Información	22
2.4.4	Trabajos Relacionados con el Área de la Recuperación de Información	22
2.5	SIMILITUD SEMÁNTICA	23
2.6	METABUSCADORES	24
	Características	25
	Ventajas	25
	Desventajas	25
2.7	SINTESIS MARCO TEORICO	26
2.8	CONCLUSIONES	27
3	DEFINICION DEL MODELO BUSQUEDA DE INFORMACION EN REDES SOCIALES EN LINEA – MOBIRSE	29
3.1	FASE DE CONCEPTUALIZACIÓN	29
3.1.1	Elementos Fundamentales Del Modelo	29
3.1.2	Restricciones Del Modelo	30

3.2	MODELO CONCEPTUAL	30
3.3	FASE DE FORMULACION	32
3.3.1	Aplicación Web RSO (App Web)	33
3.3.2	Modulo de Construcción y Actualización del Perfil de Usuario (MCAPU)	33
3.3.3	Modulo de Gestión de la Ontología (MGO)	36
3.3.4	Modulo de Expansión de Consulta (MEC)	44
3.3.5	Modulo de Reclasificación de Información (MRI)	46
3.4	FASE DE EVALUACION	47
4	IMPLEMENTACION DEL MODELO	48
4.1	FASE DE INICIO	48
4.1.1	Análisis de Requerimientos	49
4.1.2	Diagrama de Casos de Uso	49
4.1.3	Casos de Uso en Formato Compacto	50
4.2	FASE DE ELABORACIÓN	51
4.2.1	Arquitectura del Modelo	51
4.3	FASE DE CONSTRUCCIÓN	53
4.3.1	Casos de Uso en Formato Extendido	54
4.3.2	Diagramas de Secuencia	54
4.3.3	Diagramas de Clases	54
4.3.4	Diagrama de Despliegue	55
4.3.5	Modelo de Base de Datos	56
4.3.6	Desarrollo del Prototipo Software	57
5	IMPLEMENTACIÓN DE FACEBOOK META SEARCH ENGINE	60
5.1	FASE DE INICIO	61
5.1.1	Análisis de Requerimientos	62
5.1.2	Diagramas de Casos de Uso	63
5.1.3	Casos de Uso en Formato Compacto	63
5.2	FASE DE ELABORACIÓN	64
5.2.1	Arquitectura de la Aplicación	64
5.3	FASE DE CONSTRUCCIÓN	68
5.3.1	Casos de Uso en Formato Extendido	68
5.3.2	Diagramas de Secuencia	68
5.3.3	Diagrama de Clases	69
5.3.4	Diagrama de Despliegue	69
5.3.5	Modelo de Base de Datos	70
5.3.6	Desarrollo del Prototipo Software	71
5.4	FASE DE TRANSICIÓN	72
6	VALIDACIÓN DEL PROTOTIPO	74
6.1	Pruebas Alfa	74

6.1.1	CARACTERÍSTICAS DEL PROCESO DE EVALUACIÓN	74
6.1.2	PRECISIÓN	75
6.1.3	ESTADÍSTICAS KAPPA	77
6.2	Pruebas Beta	79
6.2.1	CARACTERÍSTICAS DEL PROCESO DE EVALUACIÓN	79
7	CONCLUSIONES, RECOMENDACIONES Y TRABAJO A FUTURO	82
7.1	CONCLUSIONES	82
7.2	RECOMENDACIONES	83
7.3	TRABAJO A FUTURO	83
8	REFERENCIAS BIBLIOGRÁFICAS	84

INDICE DE TABLAS

Tabla 1:	Clasificación de Ontologías	7
Tabla 2:	Metodologías para la Construcción de Ontologías	8
Tabla 3:	Herramientas para la Manipulación de Ontologías.	10
Tabla 4:	Representaciones del perfil de usuario.	13
Tabla 5:	Trabajos relacionados con Perfil de usuarios Ontológicos.	14
Tabla 6:	Clasificación de las RSO.	16
Tabla 7:	Información Básica del Perfil de Usuario en las Redes Sociales Online.	16
Tabla 8:	Estudios e Intereses de Usuario en las Redes Sociales Online.	17
Tabla 9:	Servicios que ofrecen las Redes Sociales Online.	17
Tabla 10:	Buscadores en las RSO.	19
Tabla 11:	Medidas de Similitud Semántica.	24
Tabla 12:	Conceptos de la Ontología del Perfil de Usuario	34
Tabla 13:	Casos de Uso en Formato Compacto	51
Tabla 14:	Funcionalidades API MOBIRSE	59
Tabla 15:	Plantilla para la Aplicación de MOBIRSE	59
Tabla 16:	Plantilla de Aplicación de MOBIRSE sobre FMSE	61
Tabla 17:	Casos de Uso en Formato Compacto	64
Tabla 18:	Descripción de la Tabla PostDB	71
Tabla 19:	Tiempos de Respuesta de FMSE	72
Tabla 20:	Consultas realizadas	75
Tabla 21:	Resultados de Precisión en $K = 10$	75
Tabla 22:	Apreciación de los observadores en $k=10$ resultados	78
Tabla 23:	Resultado del índice kappa en $K=10$	79

INDICE DE FIGURAS

Figura 1. Ejemplo de ontología Escritores.	6
Figura 2. Ejemplo Grafico de la Ontología Escritores.	6
Figura 3. Clasificación de las Redes Sociales.	15
Figura 4. Arquitectura Básica de un Meta buscador.	25
Figura 5. Elementos Principales para la Definición del Modelo.....	27
Figura 6. Trabajos Relacionados.....	27
Figura 7. Vista global del Modelo RI en RSO.....	30
Figura 8. Aplicación Web RSO.....	33
Figura 9. Módulo de Construcción y Actualización del Perfil de Usuario.....	35
Figura 10. Modulo de Gestión de la Ontología.....	37
Figura 11. Ejemplo de Propagación de Pesos.....	40
Figura 12. Agregar nuevo nodo, según el caso 1.....	43
Figura 13. Agregar nuevo nodo, según el caso 2.....	43
Figura 14. Agregar nuevo nodo, según el caso 2.....	43
Figura 15. Agregar nuevo nodo, según el caso 4.....	44
Figura 16. Modulo de Expansión de Consulta.....	44
Figura 17. Módulo de Reclasificación de Información.....	46
Figura 18. Diagrama de Casos de Uso.....	50
Figura 19. Arquitectura del Prototipo MOBIRSE.....	52
Figura 20. Capa Lógica del Negocio.....	53
Figura 21. Capa de Acceso a Datos.....	53
Figura 22. Diagrama de Secuencia Crear Perfil de Usuario.....	54
Figura 23. Diagrama de Clases MOBIRSE.....	55
Figura 24. Diagrama de Despliegue de MOBIRSE.....	56
Figura 25. Modelo de Base de Datos de MOBIRSE.....	57
Figura 26. Diagrama Casos de Uso Facebook Meta Search Engine.....	63
Figura 27. Arquitectura tres capas del prototipo Facebook Meta Search Engine.....	65
Figura 28. Interfaz de FMSE.....	66
Figura 29. Interfaz de FMSE tras una operación de búsqueda.....	66
Figura 30. Capa de Lógica de Negocio.....	67
Figura 31. Capa de Acceso a Datos.....	68
Figura 32. Registrar Perfil de Usuario.....	68
Figura 33. Diagrama de Clases de Facebook Meta Search Engine.....	69
Figura 34. Diagrama de Despliegue FMSE.....	70
Figura 35. Modelo de Base de Datos FMSE.....	71
Figura 36. Resultados del test de usabilidad.....	73
Figura 37. Precisión At K (K=10) para Facebook y FMSE.....	76
Figura 38. Resultados Pregunta Nro 1.....	80
Figura 39. Resultados Pregunta Nro 2.....	80
Figura 40. Resultados Pregunta Nro 3.....	81
Figura 41. Resultados Pregunta Nro 4.....	81

1 INTRODUCCION

1.1 CONTEXTO GENERAL

La continua interacción entre los seres humanos, ha dado origen a una gran variedad de relaciones entre éstos, estas relaciones varían dependiendo del contexto en que se desenvuelven los implicados, como por ejemplo en el trabajo, la universidad, la familia, entre otros. Dichas relaciones podrían conllevar a la creación de diferentes comunidades de individuos o redes sociales, en las cuales estos comparten sus sentimientos, intereses, gustos y muchos otros factores que afectan el comportamiento de cada individuo en su vida diaria.

Según Ly A [1], una red social es el conjunto de individuos o comunidades relacionados que comparten ideas y recursos en búsqueda de un objetivo común, en respuesta a una situación específica.

En las sociedades y por ende en las redes sociales, la información es un elemento que se encuentra en lo más alto de las necesidades de los seres humanos y organizaciones, lo cual es vital en el avance de éstas. El intercambio de ésta información facilita las relaciones entre los seres humanos, lo que permite generar conocimiento y en consecuencia la evolución de la sociedad misma [2-3].

Gracias a las tecnologías de la información y la comunicación (TIC), se han podido crear nuevas formas de comunicación entre las personas, lo cual les permite representar digitalmente todo tipo de comportamientos que los seres humanos experimentan en la vida real, a través del intercambio de información [4-5].

En el campo de la informática, las redes sociales son sitios web que permiten a una persona adquirir nuevas amistades y fortalecer sus lazos de amistad ya existentes, con el propósito de compartir sus preferencias, interactuar y/o crear comunidades sobre intereses comunes en distintos campos de su vida como lo son el trabajo, deporte, relaciones amorosas, arte, etc. [1].

En la actualidad las RSO ya establecidas como Facebook¹, Twitter², Myspace³, entre otras, son un punto de encuentro para millones de personas, que buscan compartir y socializar intereses [6]. Facebook, siendo la red social más popular del momento, cuenta con más de 400 millones de usuarios registrados alrededor de todo el mundo. Este sitio es el más popular para subir fotografías, con estadísticas de más de 83 millones de fotos

¹ <http://www.facebook.com/>

² <http://twitter.com/>

³ <http://www.myspace.com/>

subidas a diario. El 3 de noviembre del 2007, había siete mil (7000) aplicaciones en el sitio, cien agregadas cada día; y en enero de 2010 superaban las 500.000 [6].

Además las RSO ofrecen a sus usuarios una serie de servicios, entre los cuales se destaca el servicio de búsqueda, el cual ayuda a que dichos usuarios encuentren de una manera más fácil amigos, grupos, publicaciones y demás información alojada en estos sitios.

Teniendo en cuenta lo anterior, se puede observar que el área de la Recuperación de Información (RI) entra a jugar un papel importante en la mejora de la relevancia de la información recuperada en estos sitios web, puesto que ofrece una serie de estrategias y alternativas encaminadas a lograr este objetivo.

Por otra parte, en los últimos años, la Web semántica⁴ ha sido utilizada para mejorar los mecanismos de RI a través del uso elementos como las ontologías⁵. Las cuales han permitido que los ordenadores procesen la información de una forma más inteligente y en consecuencia los resultados obtenidos sean más precisos [7].

Otro elemento que ha resultado útil en la RI ha sido el perfil de usuario. Este ha hecho posible personalizar servicios a cada tipo de usuario y disminuir la sobrecarga de información que se le presenta, lo cual conlleva a satisfacer de una manera más amplia las necesidades de información del mismo [8].

Finalmente cabe resaltar que el uso de ontologías y el perfil de usuario aplicadas al área de la recuperación de la información sobre RSO, podrían hacer un aporte significativo a la mejora de la relevancia de la información recuperada en estos sitios web.

1.2 DECLARACION DEL PROBLEMA

De lo anterior, se puede observar que la cantidad de información existente en la Web y en especial en las redes sociales es muy grande y de un crecimiento exponencial, lo cual genera una sobrecarga de información al usuario. Por otra parte, se observa, que los buscadores de las RSO presentan algunos de los problemas existentes en los buscadores web tradicionales, como que la información recuperada no satisface las necesidades de información del usuario [9]. Además dado que los usuarios de las RSO ingresan a éstas buscando satisfacer distintas necesidades como hacer nuevas amistades, mantenerse en contacto, etc. [10], se observa que la Recuperación de la Información⁶ (RI) relevante para

⁴ La Web semántica es una extensión de la actual Web, en la cual la información se da mediante un significado bien definido, lo que facilita que los ordenadores y la gente trabajen en cooperación.

⁵ Una ontología es un vocabulario controlado que describe objetos y las relaciones entre ellos, y que tiene una gramática para usar los términos del vocabulario con el fin de expresar algo con significado dentro de un dominio de interés específico.

⁶ La recuperación de la información consiste en la recuperación automática de documentos relevantes mediante indexaciones de textos, clasificación, categorización, etc. Generalmente se utilizan palabras clave para encontrar una página relevante.

sus usuarios, se convierte en un objetivo muy deseable para las aplicaciones que funcionan sobre estos sitios [6].

En este trabajo se presenta un modelo para la recuperación de información en redes sociales Online basado en el perfil de usuario y ontologías, de tal manera que la información recuperada en estos sitios web sea personalizada, teniendo en cuenta los intereses y preferencias de los usuarios, mejorando así la relevancia de dicha información recuperada.

1.3 ESCENARIO DE MOTIVACION

Actualmente las redes sociales como Facebook y Twitter han tomado gran auge entre los usuarios de internet, puesto que se han convertido en un punto de encuentro general y de fácil acceso, lo cual conlleva a observar que los usuarios de las RSO deberían tener acceso a información relevante de una manera rápida y con el menor esfuerzo posible, para aprovechar ampliamente el potencial de la información contenida en estos sitios.

Considerando la literatura consultada, se observa que, las investigaciones realizadas sobre estos sitios web es muy poca, por ende, se dispone de un área de investigación muy amplia, y cualquier proyecto que plantee un trabajo relacionado con estos sitios, aportara un marco de referencia para futuras investigaciones en este campo.

Teniendo en cuenta el escenario anterior, y dado que el Grupo de I+D en Tecnologías de la Información – GTI, cuenta con dos áreas de investigación, a saber, web semántica y recuperación de información, se plantea el uso de estos tres elementos para generar un modelo orientado a mejorar la relevancia de la información recuperados en los sitios de redes sociales online.

1.4 CONTRIBUCIONES

- **Fortalecimiento teórico en las áreas de Recuperación de Información y Web Semántica**

A partir del estudio y el análisis de diferentes investigaciones a través de fichas bibliográficas, en adición a la presente propuesta, se fortalece la base teórica perteneciente a la institución académica en lo referente a la Web Semántica para ser aplicada a la Recuperación de Información, lo cual permite establecer un soporte para futuras investigaciones en estas áreas.

- **Modelo para la Recuperación de Información en Redes Sociales Online basado en el Perfil de Usuario y Ontologías**

La adición del factor semántico en las búsquedas Web sobre las RSO, mediante la representación del perfil de usuario con Ontologías, permiten dotar al sistema de una

mayor precisión en las búsquedas que realiza el usuario, puesto que esto permitirá personalizar los resultados de búsqueda, de acuerdo a sus intereses y preferencias.

- **Prototipo de Meta Buscador Web basado en el Modelo Propuesto**

Desarrollo de una aplicación Web acoplada a la red social Facebook, que permita mejorar la relevancia de la información recuperada, a partir de la representación del perfil de usuario con ontologías. Esta aplicación está soportada en el modelo mencionado anteriormente.

- **Artículo de Investigación**

Creación del artículo de investigación "*MOBIRSE: Un Modelo de Búsqueda de Información en Redes Sociales en Línea*", el cual fue enviado a la revista "Ingeniería e Investigación" que pertenece a la Universidad Nacional de Colombia. Este artículo se encuentra en estado de revisión por parte del personal encargado en esta revista.

- **Monografía del Trabajo de Grado**

Corresponde al presente documento, donde se describe el proceso seguido en el desarrollo del proyecto, los problemas que se presentaron, las respectivas soluciones, los principales aportes, las conclusiones y recomendaciones para el desarrollo de futuras investigaciones.

2 MARCO TEORICO

2.1 ONTOLOGIAS

Según A. Abian [7] una ontología puede definirse como un vocabulario que permite describir de forma clara objetos y las relaciones entre ellos, y que consta de una gramática mediante la cual se pueden usar los términos del vocabulario con el objetivo de expresar algo válido en un dominio en particular. Explorando el mundo de la informática se puede decir que una ontología encapsula la descripción de un área del conocimiento específica, representándola mediante un vocabulario común.

Dado que las ontologías se han convertido en un elemento importante dentro del área de la web semántica y además son un pilar fundamental dentro éste proyecto, se hace necesario revisar ciertos aspectos referentes a éstas, como son: metodologías, herramientas y aplicaciones.

2.1.1 Componentes de una Ontología

Independientemente del área del conocimiento sobre la que se pretenda trabajar, las ontologías constan de los siguientes elementos o componentes [11]:

- **Conceptos:** Son los aspectos más significativos y representativos del área del conocimiento que se quiere representar. Éstos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc. Además las clases en una ontología pueden ser organizadas en taxonomías a las que se les puede aplicar el concepto de herencia.
- **Propiedades (Atributo, Slot) [12]:** Son características que describen a la clase o concepto y por ende a sus instancias, de una manera más detallada. Dichas propiedades pueden ser representadas mediante diferentes tipos de datos, como números, cadenas de caracteres, e incluso por medio de otras clases.
- **Relaciones:** Representan el vínculo e interacción entre conceptos, generando la taxonomía del área de interés que se quiere representar. Los tipos de relaciones más comunes son: sub-clase-de, parte-de, conectada-a.
- **Instancias:** representan los objetos de determinado concepto de la ontología.
- **Axiomas:** Son los teoremas que se definen sobre las relaciones que deben cumplir los elementos que componen la ontología. Éstos deben definirse para proveer la semántica o el significado de los términos.

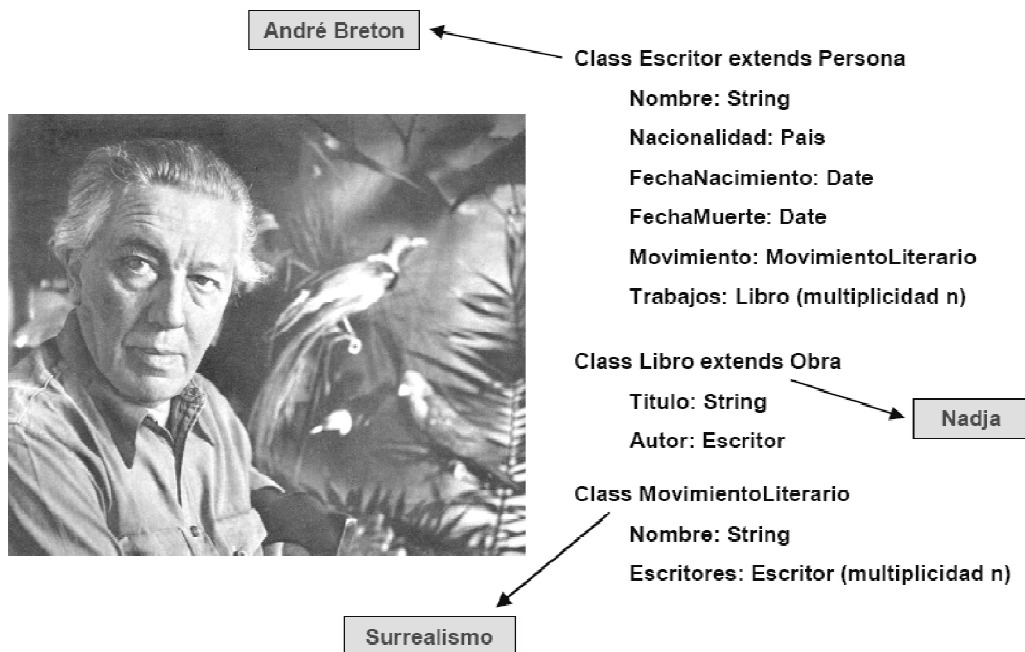
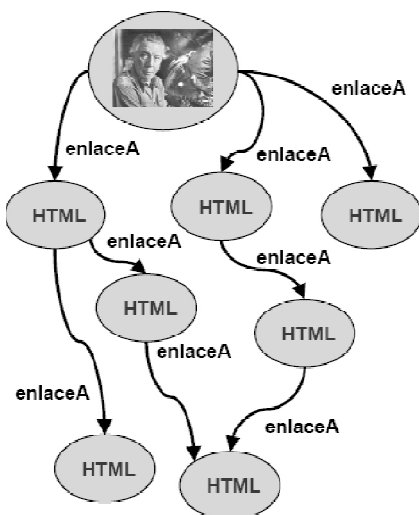


Figura 1. Ejemplo de ontología Escritores.
Fuente: El Futuro de la Web [7]

La Figura 1 muestra una ontología que está compuesta por las clases Escritor, que tiene una relación de herencia con la clase Persona y que se compone de atributos como: Nombre, Nacionalidad, etc. La clase Libro, que a su vez tiene una relación de herencia con la clase Obra y que se compone de atributos como: Titulo y Autor. Por último, la clase MovimientoLiterario, la cual tiene un Nombre y Escritores.

Red actual



Red semántica

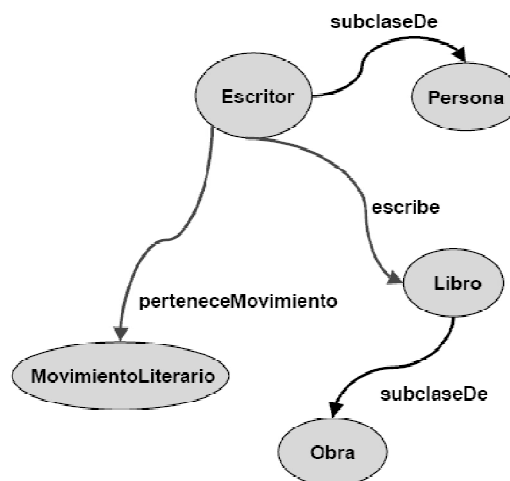


Figura 2. Ejemplo Grafico de la Ontología Escritores.
Fuente: El Futuro de la Web [7]

La Figura 2, muestra de forma grafica la ontología representada en la Figura 1. Aquí se puede observar la interacción que existe entre las clases que componen la ontología, a través de las siguientes relaciones: la clase **Escritor** es una **subclaseDePersona**, donde **subclaseDe** es la relación de herencia. La clase **Escritor** se relaciona con la clase **Libro**, a través de la relación **escribe**. Por último, la clase **Escritor** se relaciona con **MovimientoLiterario**, por medio de la relación **perteneceMovimiento**.

Además de las relaciones entre las clases de la ontología **Escritores**, la Figura 2, también muestra la diferencia entre la información representada en la red actual y la información representada en una red semántica, ya que esta última, permite realizar un proceso de inferencia sobre la información, por ejemplo, se podría determinar que André Breton, por ser **Escritor**, es una **Persona**.

2.1.2 Tipos de Ontologías

En la literatura consultada, se observa que las ontologías pueden clasificarse como se muestra en la Tabla 1[13]:

Tipo	Descripción
Ontologías de Nivel Superior	Son aquellas que permiten modelar niveles abstractos de una realidad, generando conceptos que pueden ser usados de forma genérica para clasificar términos.
Ontologías Generales	Estas permiten representar elementos como el tiempo, el espacio, eventos, entre otros. La importancia de este tipo de ontologías radica en que dichos elementos pueden ser usados en cualquier dominio.
Ontologías de Dominio	Estas ontologías sirven para representar elementos particulares de determinado ámbito, teniendo en cuenta los objetivos y las restricciones que impone el área de aplicación.

Tabla 1: Clasificación de Ontologías

2.1.3 Metodologías para la Construcción de Ontologías

Para la construcción de ontologías se hace necesario utilizar una metodología que guíe todo el proceso, el cual conlleva una serie de pasos, que ejecutados de la manera correcta dan como resultado, una representación más precisa de los elementos que queremos modelar a través de ésta. El anterior proceso es necesario ya que las ontologías se consideran un componente software en el contexto de la Web semántica[14]. Es importante destacar que la construcción de una ontología puede hacerse desde cero o reutilizar una ya existente. Estudios previos[15], han hecho un análisis comparativo de estas metodologías, teniendo en cuenta la administración del proceso del proyecto, el desarrollo de la ontología orientado a procesos y el proceso Integral. En la Tabla 2, damos una visión general de algunas de las metodologías existentes, para el desarrollo de ontologías y en la cual se muestra un cuadro comparativo

de metodologías para la construcción de ontologías. Las metodologías que se estudiaron fueron Methontology [15-16], On-To-Knowledge [15, 17], CYC [14-15, 18], USCHOLD Y KING [15, 18-19], GRÜNINGER Y FOX [15, 18-19], KACTUS [15, 18] y SENSUS [18].

Metodología	Fases			
	Inicio	Conceptualización	Implementación	Mantenimiento
Methontology	Se determina el dominio o ámbito de la ontología.	Crea un glosario de términos. Crea una taxonomía Genera un Modelo Formal.	Convertir el modelo formal a un modelo computable utilizando un lenguaje para construir ontologías.	Se verifica el funcionamiento de la ontología.
On-To-Knowledge	Se especifica el objetivo de la ontología y su alcance.	Genera un Modelo Formal.	Se genera un modelo computable a partir del Modelo Formal.	Se verifica el funcionamiento de la ontología y se define el responsable de su mantenimiento.
CYC			Codificación manual de conocimiento Codificación de conocimiento asistido por herramientas software. Codificación de conocimiento realizado por herramientas software	
USCHOLD Y KING	Identificar el propósito de la ontología.	Se identifican conceptos y sus relaciones fin de generar un Modelo Formal.	Codificación de la ontología utilizando un lenguaje formal para ontologías.	Se verifica el funcionamiento de la ontología y se documenta todo lo referente a ésta.
GRÜNINGER Y FOX	Determinar posibles usos y alcance de la ontología.	Se identifican conceptos, sus relaciones y se expresan en lógica de primer orden.		
KACTUS	Se define el contexto o dominio de la ontología y se identifican conceptos y sus relaciones.	Se realiza un diseño preliminar de la ontología. Se descompone la ontología para reestructurarla a fin de lograr un diseño modular.		
SENSUS		Se seleccionan los términos (semilla) más relevantes pertenecientes al dominio de la ontología.	Se enlazan los términos (semilla) seleccionados a la ontología SENSUS. Los conceptos que se encuentran en la ruta que va desde los términos (semilla) hasta la raíz de SENSUS son incluidos. Finalmente, términos que podrían ser relevantes para el dominio y que aun no han aparecido son agregados.	

Tabla 2. Metodologías para la Construcción de Ontologías

A parte de las metodologías mostradas en la Tabla 2, existen otras metodologías para la construcción de ontologías basadas en escenarios, un claro ejemplo de esto lo muestra la metodología Neón⁷, la cual, provee una guía clara para abordar todos los aspectos claves, relacionados con el proceso ingenieril de una ontología. Los nueve escenarios propuestos en la metodología, cubren las situaciones que comúnmente se presentan en el desarrollo de una ontología. Estos son:

- De la especificación a la implementación.
- Reutilización y reingeniería de recursos NO ontológicos.
- Reutilización de recursos ontológicos.
- Reutilización y reingeniería de recursos ontológicos.
- Reutilización y combinación de recursos ontológicos.
- Reutilización combinación y reingeniería de recursos ontológicos.
- Reutilización de diseño de patrones de ontologías.

⁷http://www.neon-project.org/hw/NeOn_Book

- Reestructuración de recursos ontológicos.
- Localización de recursos ontológicos.

De lo anterior, se puede observar, que pocas metodologías abarcan las cuatro fases generales para la construcción de una ontología, ahora bien, dichas metodologías son las más indicadas para construir una ontología desde el punto de vista de la ingeniería de software, razón por la cual, a nuestro criterio, la metodología más adecuada para llevar a cabo el proceso de construcción de la ontología sería Methontology, puesto que sus fases son claras y bien definidas, además esta metodología guía el proceso de construcción de la ontología como un componente software.

2.1.4 Herramientas para el Desarrollo de Ontologías

Para la construcción de ontologías, además de la metodología, se debe tener en cuenta la herramienta o el lenguaje a utilizar. En la actualidad existen muchas herramientas útiles para la construcción y edición de ontologías. En la Tabla 3, se muestran las más importantes herramientas que existen actualmente:

Herramientas para Desarrollar Ontologías				
Nombre	Tipo	Entorno Colaborativo	Servicios	Otras Características
Protégé [20]	Open Source	Si	- Creación. - Edición. - Visualización. - Manipulación.	Modelado a través de Marcos y OWL. Soporte para múltiples formatos.
Ontolingua ⁸	Open Source	Si	- Análisis. - Creación. - Edición. - Intercambio.	Proporciona una biblioteca de ontologías modulares y reutilizables.
WebOnto ⁹	Open Source	Si	- Visualizar. - Crear. - Editar.	Facilita a los usuarios la ampliación de ontologías de gran tamaño.
Neon ¹⁰	Open Source	Si	- Creación. - Mantenimiento.	Ofrece un plugin para cada actividad especificada en la metodología de desarrollo Neón.
Kaon ¹¹	Open	No	- Creación.	Ofrece otro tipo de

⁸<http://www.ksl.stanford.edu/software/ontolingua/>

⁹<http://projects.kmi.open.ac.uk/webonto/>

¹⁰http://neon-toolkit.org/wiki/Main_Page

¹¹http://kaon.semanticweb.org/main_kaonOverview.pdf/view

	Source		<ul style="list-style-type: none"> - Almacenamiento. - Recuperación. - Mantenimiento. - Aplicación. 	herramientas para construir aplicaciones basadas en ontologías.
OntoStudio ¹²	Comercial	Si	<ul style="list-style-type: none"> - Modelado. - Creación. - Mantenimiento. - Herramientas de mapeo. - Editor grafico de reglas. - Entorno integrado de pruebas. 	Soporte para varios formatos.

Tabla 3. Herramientas para la Manipulación de Ontologías.

Teniendo en cuenta la comparación de la Tabla 3, la mayoría de las herramientas analizadas, son útiles para la manipulación de una ontología, sin embargo, Protégé, siendo la herramienta más utilizada por los desarrolladores de ontologías, goza de mayor documentación, soporte, interfaz amigable para el usuario y es software libre. Por las razones nombradas anteriormente, esta herramienta será la seleccionada, para abordar nuestro proyecto.

2.2 PERFILES DE USUARIO

Se puede considerar el perfil de usuario como una colección de información personal, la cual puede incluir información demográfica como: nombre, edad, país, nivel de educación, etc. Dicho perfil permite representar intereses o preferencias de un grupo de usuarios o de una sola persona [21-22].

La importancia del perfil de usuario radica en la personalización de servicios (comercio electrónico, aprendizaje electrónico, sistemas de recomendación, etc.) y de información (correo electrónico, noticias, grupos de noticias, documentos web, etc.). Igualmente el perfil de usuario también puede ser utilizado para mejorar la navegación, permitiendo crear sitios web adaptativos.

Como se puede observar el perfil de usuario, es un elemento útil que va orientado a disminuir información no relevante, lo cual provoca un impacto favorable en el usuario, puesto que éste obtiene lo que realmente necesita sin realizar mucho esfuerzo.

2.2.1 Representación del Perfil de Usuario

¹²<http://www.ontoprise.de/en/home/products/ontostudio/>

Existen múltiples formas de representar el perfil de usuario, entre las más utilizadas se encuentran, la representación por medio de palabras claves ponderadas, redes semánticas, conceptos ponderados y reglas de asociación [22].

2.2.1.1 Perfil de Usuario basado en Palabras Claves

La representación de perfiles de usuario por medio de palabras claves ponderadas es la más común. Estas palabras clave, pueden ser extraídas de manera automática (durante la navegación del usuario, páginas Web marcadas o salvadas por el usuario) o proporcionadas por el usuario. Las ponderaciones o pesos son representaciones numéricas asociadas a las palabras claves, que representan intereses del usuario, los cuales son utilizados para representar la importancia de ese tema de interés. Cada una de las palabras claves puede referenciar a un tema de interés del usuario, ahora bien, muchas palabras claves agrupadas pueden representar categorías de intereses de los usuarios, lo cual conlleva a que se puedan generar estándares de preferencias.

2.2.1.2 Perfiles de usuario basado en Redes Semánticas

Esta alternativa surge a razón de combatir los problemas de polisemia¹³ que se presentan en la construcción de perfiles de usuario basados en palabras claves. Estos perfiles son representados por una red semántica ponderada, donde cada nodo representa un concepto. Muchos proyectos han basado su sistema de recuperación de información en este tipo de técnica, por ejemplo, el proyecto InfoWeb [23], hace uso de un sistema de redes semánticas para representar los intereses del usuario a largo plazo, modelando el perfil de usuario como una red semántica de conceptos.

2.2.1.3 Perfiles de usuario basados en Conceptos

Los perfiles construidos bajo esta técnica, son similares a los perfiles de usuario construidos con redes semánticas, puesto que ambas técnicas se basan en nodos y relaciones entre ellos. A diferencia de las redes semánticas, los nodos representan temas abstractos, en lugar de palabras específicas o conjuntos de palabras relacionadas. Este tipo de perfiles también tienen cierta similitud con los perfiles construidos con palabras clave, ya que, estos pueden representarse como vectores de características ponderadas, solo que, dichas características representan conceptos¹⁴ en lugar de palabras o conjunto de palabras clave¹⁵. Diferentes mecanismos son utilizados, para determinar qué nivel de

¹³ Polisemia: en lingüística se presenta cuando una misma palabra o signo lingüístico tiene varios significados.

¹⁴ PU basado en Conceptos: representan temas abstractos que se consideran interesantes para el usuario.

¹⁵ PU basado en Palabras Clave: representan temas abstractos que se consideran interesantes para el usuario.

interés muestra el usuario para determinado tema, la técnica más simple es asignar un peso numérico a cada tema.

2.2.1.4 Perfiles de usuarios basados en Ontologías

Los perfiles basados en ontologías o perfiles de usuario ontológicos, son utilizados para representar a un usuario como una instancia de una ontología de dominio, como por ejemplo, el trabajo realizado por A. Sieg et al [8], propone modelar el contexto de usuario a través de un perfil de usuario ontológico, para mejorar la relevancia de los resultados obtenidos en la búsqueda Web. Esto es posible, asignando puntajes a conceptos existentes en una ontología de dominio, los cuales inicialmente toman el valor de uno. A medida que el usuario interactúa con el sistema, estos valores son actualizados y gracias al mecanismo de propagación, los conceptos son activados o desactivados según el puntaje que poseen.

La Tabla 4 muestra una síntesis de las representaciones de perfil de usuario explicadas hasta este punto.

Tipo	Fuentes de información	Técnicas de construcción	Ejemplos
Perfil de Usuario basado en Palabras Claves	Páginas Web	<ul style="list-style-type: none"> - Extracción de palabras clave de mayor puntaje. - Vectores de documentos - Comparación de vectores de intereses. - Mezcla de vectores de intereses más cercanos. 	Amalthea [24], WebMate [25], Alipes [26]
Perfiles de usuario basado en Redes Semánticas	<ul style="list-style-type: none"> - Documentos de muestra - Colección de documentos de estereotipo - Páginas Web - Intervención del usuario 	<ul style="list-style-type: none"> - Extracción de palabras clave de mayor puntaje. - Creación de un nodo por palabra. - Enlazar nodos cuyas palabras concurren en varios documentos. - Mapeo de palabras a conceptos usando WordNet - Aprendizaje de conceptos usando redes neuronales. - Creación de nodo por concepto. - Creación de nodo por concepto y por palabra con ayuda del experto humano. 	ifWeb [27], SiteF [28], PIN [29], InfoWeb [23], WIFS [30]
Perfiles de usuario basados en Conceptos	<ul style="list-style-type: none"> - Páginas Web pre-clasificadas. - Resultados de búsqueda. - Artículos de investigación pre-clasificados. - Manipulación del perfil. 	<ul style="list-style-type: none"> - Coloreo de arboles. - Agrupamiento. - Propagación de conceptos padres. - Clasificación de texto e identificación de conceptos. - Expansión del clasificador de entrenamiento, basado 	Persona [31], ARCH [32], Foxtrot [33], OBIWAN [34], Mismatch [8], Liu et al [35], PVA [36], Bibster [37]

		en retroalimentación. - Adaptar taxonomía para añadir o remover conceptos. - Manipulación directa.	
--	--	--	--

Tabla 4. Representaciones del perfil de usuario.

2.2.2 Perfiles de usuarios basados en Ontologías

A continuación, se describen las principales características de algunos proyectos que han hecho uso de perfiles de usuario basados en ontologías:

2.2.2.1 Creación de una Ontología para el Perfil de Usuario: Método y Aplicaciones

En esta investigación se propone una ontología general y adaptable, que pueda ser utilizada por cualquier aplicación de acuerdo a sus características. Dicha ontología usa el enfoque top-down, es decir, de lo general a lo específico y se basa en las características estáticas del usuario, es decir, intereses a largo plazo, además puede ser extendida por medio de herencia o adición de clases [38].

2.2.2.2 Personalización de Búsqueda Web con Perfiles de Usuario Ontológicos

Esta investigación se centra en modelar el contexto de usuario, haciendo uso de tres elementos, a saber: la necesidad del usuario a corto plazo, el conocimiento semántico de un dominio específico y los intereses a largo plazo. La idea es asignar puntajes a los intereses derivados de los conceptos existentes en la ontología de dominio, con lo cual se puede definir una jerarquía de temas que puede ser de interés para el usuario.

La actualización del perfil de usuario, se logra gracias al mecanismo de activación de propagación y la recolección de la información para construir y/o mantener el perfil de usuario, se hace de manera implícita, además, los conceptos son ponderados, es decir, que tienen un peso asignado inicialmente de uno [8].

2.2.2.3 Búsqueda Personalizada Basada en Sesión Usando un Perfil de Usuario Ontológico

En este proyecto el perfil de usuario, es representado como un grafo de los conceptos más relevantes de la ontología, en la misma sesión de búsqueda y no como una instancia de la ontología entera. Para la construcción de este perfil, las principales fases fueron:

- Construir el perfil de búsqueda en base a las actividades de búsqueda.
- Inicializar y mantener el perfil de búsqueda, sobre la sesión de búsqueda.
- Detectar un nuevo límite de sesión, cuando una nueva consulta es realizada.

Para la construcción y actualización del perfil de usuario, se tuvieron en cuenta los intereses a corto plazo, conceptos ponderados y un mecanismo de límite de sesión, el cual es utilizado para determinar si una nueva consulta se ajusta o no dentro del perfil actual [39].

2.2.2.4 Perfilamiento de Usuario Ontológico sobre Recomendación Personalizada en Comercio Electrónico

Este trabajo presenta un novedoso modelo de recomendación, que involucra el manejo de ontologías para representar los intereses en una comunidad de comercio electrónico. Lo anterior, con el objetivo, de mejorar los ítems o productos recomendados en base al puntaje o etiquetado social que los usuarios dan sobre determinados productos. Cabe resaltar que con base en dicho etiquetado, se calculan perfiles diarios para los usuarios y se sugieren productos que el usuario no ha visto, pero que resultan ser atractivos por ser los de mayor puntaje y popularidad dentro de la comunidad. Entre las técnicas utilizadas para llevar a cabo este proyecto, se destacan [40]:

- Similitud semántica entre conceptos.
- Inferencia hacia arriba.
- Inferencia hacia abajo.

La Tabla 5 muestra a modo de resumen los trabajos realizados sobre perfil de usuario.

Elemento	M. Golemati	A. Sieg	M. Daoud	Siping He
Intereses a corto plazo.		✓	✓	✓
Intereses a largo plazo.	✓	✓		✓
Perfil representado con ontología	✓	✓		✓
Perfil representado con grafo			✓	
Uso de conceptos ponderados		✓	✓	

Tabla 5. Trabajos relacionados con Perfil de usuarios Ontológicos.

Según muestra la Tabla 5, las ontologías son un elemento muy deseable, en el momento de representar perfiles de usuario, ya que estos elementos, permiten inferir conocimiento acerca de los intereses y necesidades del usuario. Ahora bien, para que la representación del perfil usuario sea más precisa, se hace necesario tener en cuenta algunas características como los intereses a largo y corto plazo, etc. Para abordar el presente proyecto, sería ideal representar el Perfil de Usuario por medio de una ontología y que sus conceptos sean ponderados.

2.3 REDES SOCIALES

Se puede decir, que una red social es el conjunto de individuos o comunidades relacionados que comparten ideas y recursos en búsqueda de un objetivo común, en respuesta a una situación específica. En el campo de la informática, las Redes Sociales Online (RSO), son sitios web que permiten a una persona adquirir nuevas amistades y fortalecer sus lazos de amistad ya existentes, con el propósito de compartir sus preferencias, interactuar y/o crear comunidades sobre intereses comunes en distintos campos de su vida como lo son el trabajo, deporte, relaciones amorosas, arte, etc. [1].

2.3.1 Tipos de Redes Sociales

Actualmente las redes sociales pueden clasificarse en dos grandes grupos: las redes sociales horizontales y las verticales [41], como se observa en la Figura 3.

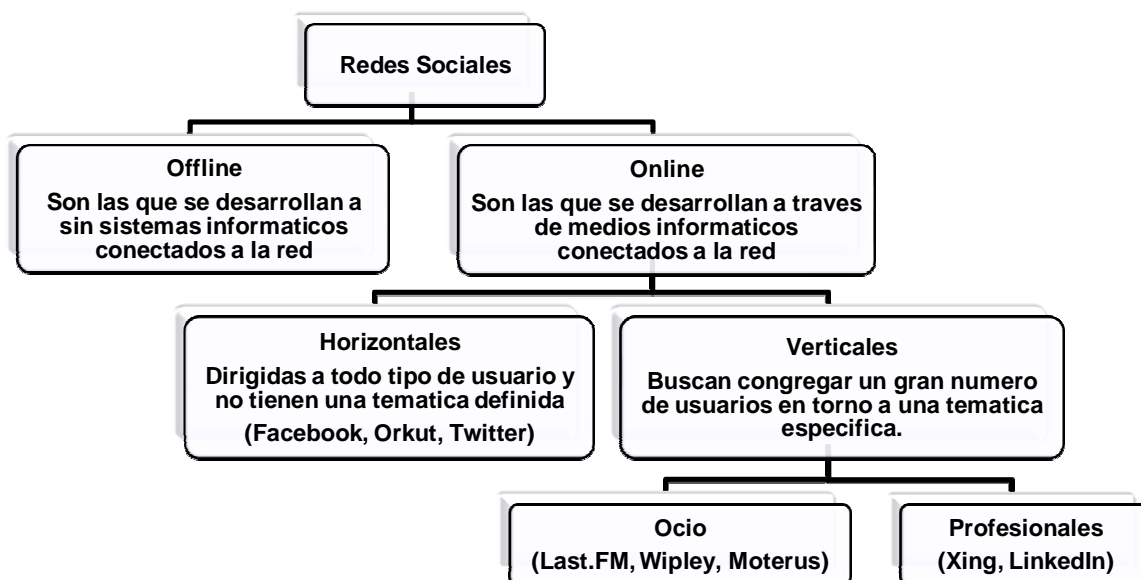


Figura 3. Clasificación de las Redes Sociales.

2.3.2 Características de las Redes Sociales

Dada la gran variedad y tipo de RSO existentes, es necesario revisar las características de cada una de estas, teniendo en cuenta los datos que manejan y los servicios que ofrecen. Puesto que, en la literatura existente, no se encontró una base teórica, que muestre esta información, se optó por realizar una revisión general teniendo en cuenta las RSO más populares actualmente¹⁶. Según la Figura 3, las RSO estudiadas, se pueden clasificar como se muestra en la Tabla 6.

¹⁶ <http://www.ebizmba.com/articles/social-networking-websites>

RSO	TIPO
Facebook [42]	Horizontal
Myspace [43]	Horizontal
Twitter [44]	Horizontal
LinkedIn [45]	Vertical Profesional
Last.fm [46]	Vertical Ocio
Tagged [47]	Horizontal
Hi5 [48]	Horizontal
Meetup [49]	Vertical Profesional
Bebo [50]	Horizontal
Mylife [51]	Vertical Ocio
Friendster [52]	Horizontal
Myheritage [53]	Vertical Ocio
Multiply [54]	Horizontal

Tabla 6. Clasificación de las RSO.

En la Tabla 7, se muestran los datos que hacen parte de la información básica del perfil del usuario, que manejan las distintas RSO:

Datos del perfil de usuario en las Redes Sociales Online													
Atributo	Facebook	Myspace	Twitter	LinkedIn	Last.fm	Tagged	Hi5	Meetup	Bebo	Mylife	Friendster	Myheritage	Multiply
Información Básica													
Foto perfil	X	X	X	X	X	X	X	X	X	X	X	X	X
Alias	X	X	X	X	X	X	X	X	X	X	X	X	X
Nombres	X	X	X	X	X	X	X	X	X	X	X	X	X
Apellidos	X	X	X	X	X	X	X	X	X	X	X	X	X
Fecha Nacimiento	X	X	X	X	X	X	X		X	X	X	X	
Sexo	X	X		X	X	X	X		X	X	X	X	
Relación familiar	X				X							X	
Relación sentimental	X	X		X		X	X		X	X	X		X
Ciudad de origen	X	X					X		X		X	X	X
Ciudad actual	X		X			X		X	X	X	X	X	
Creencias Religiosas	X	X				X	X		X	X			X
Biografía	X	X	X		X	X	X	X	X		X	X	X
Citas favoritas	X						X		X			X	
Ideología política	X									X		X	
Email	X	X	X	X	X	X	X	X	X	X	X	X	X
Mensajería instantánea	X			X			X		X				X
Teléfono	X			X			X		X	X		X	X
Dirección	X			X			X		X	X	X	X	X
Cod. postal	X			X					X	X		X	
Vecindario	X												
Sitio Web			X	X	X					X		X	X
Descripción física		X								X		X	
Estilo de vida.		X								X			X
Etnia		X				X	X			X			
Orientación Sexual		X				X				X			

Tabla 7. Información Básica del Perfil de Usuario en las Redes Sociales Online.

De la Tabla 7, se observa que, datos como nombres, apellidos, fecha de nacimiento, email y biografía, están presentes en todas las RSO analizadas, lo cual lleva a considerar

que dichos datos, deben estar presentes en cualquier representación del perfil de usuario que se realice. Para complementar esta información, la Tabla 8 muestra otros ítems que pueden hacer parte del perfil de usuario.

Atributo	Facebook	Myspace	Twitter	LinkedIn	Last.fm	Tagged	Hi5	Meetup	Bebo	Mylife	Friendster	Myheritage	Multiply
Estudios													
Estudios de Secundaria	X	X							X	X	X	X	X
Estudios Universidad	X			X		X			X	X	X	X	X
Empleo	X	X		X		X			X	X	X	X	X
Intereses													
Actividades	X	X		X			X					X	X
Hobbies	X									X		X	X
Música	X	X			X		X		X	X	X	X	
Libros	X	X					X		X	X	X	X	
Cine	X	X					X		X	X	X	X	
Televisión	X	X					X		X	X	X	X	
Páginas	X												
Héroes		X											
Deportes									X	X		X	
Destinos										X		X	
Restaurante												X	
Celebridades												X	
Comida												X	
Busco (Interés sentimental)	X						X			X	X		X

Tabla 8. Estudios e Intereses de Usuarios en las Redes Sociales Online.

De la Tabla 8, se observa, que las RSO horizontales, por ser de temática abierta, manejan mucha más información del perfil de usuario que las RSO verticales, lo cual muestra, que de estas se pueda extraer mucha más información de los usuarios y en consecuencia sean las más favorables para desarrollar aplicaciones para estos sitios.

Servicios	Facebook	Myspace	Twitter	LinkedIn	Last.fm	Tagged	Hi5	Meetup	Bebo	Mylife	Friendster	Myheritage	Multiply
Servicios													
Chat	X	X											X
Mensajes	X	X	X	X	X	X	X		X	X	X		X
Eventos	X	X			X					X		X	
Contactos	X	X	X	X	X	X	X	X	X	X	X	X	X
Grupos	X	X		X	X	X	X		X		X		X
Aplicaciones	X	X		X	X	X	X	X	X		X	X	
Publicaciones	X	X	X	X	X	X	X		X		X	X	X
Páginas	X											X	
Etiquetas	X	X			X		X		X		X	X	X
API	X		X	X	X			X	X		X		X
Foros	X	X			X	X			X		X	X	
Blog			X	X	X	X		X	X		X	X	X
Buscador	X	X	X	X	X	X	X	X	X	X	X	X	X
Tipo de Información													
Fotos	X	X	X	X	X	X	X	X	X	X	X	X	X
Videos	X	X				X	X		X			X	X
Enlaces	X	X	X	X	X	X	X	X	X	X	X	X	X
Texto	X	X	X	X	X	X	X	X	X	X	X	X	X

Tabla 9. Servicios que ofrecen las Redes Sociales Online.

De la Tabla 9 se puede concluir lo siguiente:

- Las RSO que ofrecen gran variedad de servicios como Facebook o Myspace, son las más visitadas, lo cual nos lleva a concluir que debido a ese criterio, es que estas RSO gozan de mayor popularidad.
- Se observa que en todas las RSO, sobre las cuales se realizó este estudio, existen servicios en común, tales como: mensajería, contactos, publicaciones y

buscadores, lo cual nos lleva a considerar, que estos servicios son los más aceptados por los usuarios.

- Se observó, que el tipo de información que todas estas RSO manejan son fotos, enlaces y texto.
- Se observa, que servicios como el API de desarrollo, son escasos en la mayoría de RSO analizadas, lo cual reduce nuestras posibilidades, en el momento de elegir la RSO que será utilizada como caso de estudio, para probar el modelo planteado.
- Se puede observar que la mayoría de las RSO de este estudio, pertenecen a la categoría de Redes Sociales Horizontales, demostrando que estas son las más utilizadas por los usuarios de Internet.
- Las RSO horizontales, por ser de temática abierta, manejan mucha más información del perfil de usuario que las RSO verticales, tal como se ve en la sección de intereses de la tabla anterior, lo cual conlleva a que de estas se pueda extraer mucha más información de los usuarios.
- Las RSO horizontales, es decir, las de temática abierta, son las que gozan de mayor popularidad entre los usuarios, según lo indican ebizmba¹⁷, toptenrewies¹⁸, webtrends¹⁹ y metrolic²⁰.

Entre los servicios que más se destacan y en el que más énfasis se realizara en este estudio, es en los buscadores, esto debido, a que la mejora de la relevancia en la información recuperada, se ha convertido en un objetivo más para los desarrolladores de estos sitios. En la Tabla 10 se muestran algunas de las características más importantes de los buscadores en las RSO actuales.

RSO	Características				
	Agrupar	Predefine tema	Asociado a otro buscador	Autocompletado	Búsqueda avanzada
Facebook	✓		✓	✓	✓
MySpace		✓	✓		✓
Twitter	✓				
LinkedIn		✓		✓	✓
Last.fm	✓			✓	
Tagged		✓			✓
Hi5	✓			✓	✓
Meetup		✓			
Bebo	✓				✓
Mylife		✓			

¹⁷<http://www.ebizmba.com/articles/social-networking-websites>

¹⁸<http://social-networking-websites-review.toptenreviews.com/>

¹⁹<http://webtrends.about.com/b/2010/03/15/the-top-10-most-popular-social-networks.htm>

²⁰<http://www.metrolic.com/top-10-social-networking-sites-137726/>

Friendster		✓			
Myheritage	✓				✓
Multiply		✓	✓		

Tabla 10. Buscadores en las RSO.

De la Tabla 10, se puede observar que la RSO que ofrece un servicio de búsqueda más completo, es Facebook. Ahora bien, en cuanto a la precisión de los resultados de búsqueda, estos están dados de acuerdo al tipo de RSO, es así como por ejemplo, en Facebook, la búsqueda de empleos es menos precisa, que si se realiza en LinkedIn. Otro aspecto que vale la pena resaltar, es el autocompletado, ya que por ejemplo en Last.fm, Hi5 y LinkedIn, este servicio no solo da sugerencias, si no que las agrupa, según el tipo de información que maneje la RSO. Finalmente vale la pena aclarar, que mientras unos buscadores permiten una búsqueda general y luego agrupan los resultados, otros definen un buscador para cada tema que maneja la RSO (contactos, grupos, fotos, etc.).

Un elemento a tener en cuenta en las RSO, es si estas ofrecen una API para los desarrolladores, ya que a través de estas, distintas aplicaciones (Web, escritorio o móvil) pueden interactuar con la RSO. Según la exploración realizada, las RSO con una API de desarrollo son: Facebook, Twitter, LinkedIn, Last.fm, Meetup, Bebo, Friendster y multiply. De las anteriores RSO, es importante resaltar, que las API que gozan de mayor popularidad, son las de Facebook y Twitter. Lo cual indica, que si se desea crear una aplicación para una RSO, estas dos serian las más adecuadas.

2.3.3 Uso y motivaciones en las RSO

Dada la importancia que las RSO han adquirido en los últimos años, se hace importante conocer las motivaciones por las cuales estos sitios son utilizados por los usuarios de Internet. Esto es importante, ya que al tener en cuenta estos aspectos, se podrán ofrecer servicios e información que satisfagan de mejor manera a los usuarios.

Por ejemplo, en la RSO Facebook, el estudio hecho por Adam N. Joinson [10], ha identificado las siguientes motivaciones de uso de esta importante RSO.

- Conexión Social: Descubrir que están haciendo ahora los viejos amigos, volver a contactar con personas que ha perdido el contacto, aceptar solicitudes de amistad, encontrar personas que no han visto en un tiempo, mantener contacto con personas que no pueden ver muy a menudo, ponerse en contacto con amigos que están lejos de casa.
- Intercambio de identidades: Organizar o unirse a eventos, unirse a grupos, comunicarse con personas que tienen ideas afines.
- Fotografías: Mirar fotos, ser etiquetado en fotos, etiquetar fotos, intercambiar o subir fotografías.
- Contenido: Aplicaciones dentro de Facebook, jugar, descubrir aplicaciones que han sido agregadas por los amigos, quizzes.
- Investigación Social: Observar virtualmente a la gente, utilizar la búsqueda avanzada para mirar tipos específicos de personas, conocer gente nueva, espiar personas.

- Navegar en las Redes Sociales: Mirar los perfiles de las personas que no conoce, ver los amigos de otras personas, navegar por los amigos de sus amigos.
- Actualización del estado: Actualización de su propio estado, publicar sus pensamientos, mirar los estados de las otras personas.

2.4 RECUPERACION DE LA INFORMACION

La Recuperación de la Información (RI), es el área de la informática, que se encarga de encontrar recursos (texto, videos, imágenes, etc.), que satisfagan las necesidades de información de los usuarios [55].

2.4.1 Modelos de Recuperación de información

Para la RI existen tres modelos clásicos que se explicaran a continuación:

2.4.1.1 Modelo Booleano

Este modelo para la RI, se basa en el algebra de Bool y la teoría de conjuntos, donde las consultas son planteadas en forma de expresión booleana, es decir, los términos de las consultas son combinadas con los operadores AND, OR y NOT. Además, este modelo ve cada documento como un conjunto simple de palabras [55] y se considera como el fácil de implementar [56].

2.4.1.2 Modelo Probabilístico

Este modelo también es conocido como Modelo de Recuperación de Independencia Binaria (BIR). La idea es trabajar sobre los términos que conforman la consulta, asignándole un peso a cada uno de ellos, según ayuden a filtrar los documentos relevantes de los irrelevantes. Teniendo en cuenta lo anterior, este modelo puede calcular el grado de similitud que existe entre cada documento recuperado y la consulta ponderada, ordenando los documentos de mayor a menor, según su probabilidad de relevancia [56].

2.4.1.3 Modelo Vectorial

A diferencia de los modelos explicados anteriormente, éste tiene en cuenta la frecuencia con la que aparecen los términos de la consulta en los documentos recuperados, lo cual ayuda a identificar, la importancia de un término en determinado documento. Finalmente, se puede decir que el modelo vectorial, funciona con base en tres principios fundamentales, a saber: La equiparación parcial, la ponderación de los términos en los documentos y la ponderación de los términos en la consulta [56].

2.4.2 Técnicas para la Recuperación de la Información.

Existen varias técnicas que se utilizan para la RI, a continuación se nombraran algunas de las nombradas:

2.4.2.1 Sistemas de Recuperación de Logia Difusa

Esta técnica permite la recuperación de documentos similares a los solicitados en la consulta, sin que necesariamente, los términos de la consulta aparezcan en los documentos, es decir, recuperar documentos que aborden la misma temática, sin que estos contengan los términos de la consulta [57].

2.4.2.2 Técnica de Ponderación de Términos

Esta técnica intenta asignar un valor o peso específico a los documentos recuperados, según los intereses del usuario. Dicho valor depende de los términos que el documento contiene y la frecuencia con que se repite. Haciendo que, el documento más relevante para el usuario, es aquel contenga todos los términos de búsqueda y que el valor de frecuencia sea el más alto [58].

2.4.2.3 Técnicas de Clustering

Esta técnica busca agrupar documentos, en subgrupos o clusters. Dichos grupos deben ser internamente similares, pero claramente diferentes de los otros subgrupos [55]. Para la aplicación de esta técnica, se pueden utilizar distintos algoritmos, algunos de ellos son [58]:

- Algoritmo K-means
- COBWEB
- Algoritmo EM

2.4.2.4 Técnicas de retroalimentación por Relevancia

La idea de esta técnica es involucrar al usuario en el proceso de recuperación de la información, con el objetivo de mejorar los resultados finales. Para llevar esto a cabo, se realiza el siguiente proceso [55]:

- El usuario realiza una consulta.
- El sistema retorna un conjunto inicial de resultados.
- El usuario marca los resultados como relevantes y no relevantes.
- El sistema calcula una mejor representación de las necesidades de información, basado en la retroalimentación hecha por el usuario.
- El sistema muestra una serie de resultados ya revisados.

2.4.2.5 Técnica de Stemming

Esta técnica pretende eliminar las confusiones semánticas, originadas por los prefijos y los sufijos a la hora de buscar conceptos, para esto solo tiene en cuenta la raíz de la palabra. Con el objetivo de separar la raíz de los sufijos y prefijos, existen algoritmos que realizan este truncamiento, algunos de ellos son [58]:

- Paice/Husk
- S-stemmer / n-gramas
- Técnicas lingüísticas

2.4.3 Evaluación en la Recuperación de la Información

Existen algunos índices o medidas, que son utilizados para evaluar la calidad de la información recuperada, a continuación se nombran algunas de las más usadas [58]:

- **Exhaustividad:** consiste en recuperar todos los documentos que puedan ser relevantes de una colección, de acuerdo a los parámetros sugeridos por el usuario.
- **Precisión:** Esta medida representa el porcentaje de los documentos recuperados, que son relevantes de acuerdo a determinada consulta. Se calcula así: documentos relevantes recuperados, sobre, total de documentos recuperados.
- **Relevancia:** Mide la importancia que el documento recuperado representa para el usuario, según sus necesidades de información.

Existen otras medidas de evaluación que están ligadas a la satisfacción del usuario, siendo estas más difíciles de evaluar, por esta razón, en algunos casos están dejando de ser utilizadas, a continuación se nombran algunas de ellas [59]:

- Ratio de cobertura.
- Ratio de novedad.
- Exhaustividad relativa.

2.4.4 Trabajos Relacionados con el Área de la Recuperación de Información

2.4.4.1 Meta Buscador Web Semántico Basado en una Taxonomía, Ontologías y Perfil de Usuario

En este proyecto se define un modelo de un meta buscador web semántico, a fin de filtrar y reordenar documentos web retornados originalmente por buscadores tradicionales. Esto se logra integrando una taxonomía general de conocimiento, una ontología de dominio general, ontologías específicas y el perfil de usuario, con el objetivo de mejorar la precisión de los resultados de búsqueda desplegados al usuario [60].

2.4.4.2 Modelo de un meta buscador web basado en taxonomías, ontologías y retroalimentación del usuario

Este proyecto propone un modelo de un meta buscador que utiliza taxonomías generales de conocimiento y ontologías para encontrar la similitud semántica que existe entre los documentos retornados por los buscadores y la consulta hecha por el usuario. El modelo usa las ontologías para expandir la consulta que se envía a los buscadores tradicionales y además aprovecha la retroalimentación que el usuario da al sistema (basado en una calificación de los resultados presentados), buscando con ello mejorar la relevancia de los resultados de búsqueda [61].

2.5 SIMILITUD SEMÁNTICA

La similitud semántica, se refiere al grado de proximidad entre dos conceptos de una ontología dada [62], es decir, se refiere a que tan relacionados están estos conceptos entre sí, ahora bien, para determinar el grado de relación entre dichos conceptos, se hace necesario utilizar una medida que nos proporcione de manera cuantificable esta relación. En la Tabla 11, se muestran algunos tipos de medidas de similitud semántica. Cabe aclarar, que las siguientes medidas solo son aplicables a ontologías de tipo jerárquico (como por ejemplo WordNet y ODP), a excepción de las medidas de similitud presentadas en [62-63].

Medida de Similitud Semántica	Descripción
$dist_{rad}(c_1, c_2) = len(c_1, c_2)$	Esta medida propuesta por Rada [64], representa la similitud semántica como la menor distancia semántica entre dos conceptos, donde <i>len</i> representa la longitud de la ruta más corta que une los dos conceptos en la jerarquía, teniendo en cuenta que la distancia se mide según el número de enlaces.
$sim_{res}(c_1, c_2) = CI(pcc(c_1, c_2))$	Esta medida propuesta por Resnik [65], asocia un peso a cada nodo, donde dicho peso representa el contenido de información (<i>CI</i>) del concepto, es decir, los pesos más generales de la ontología tienen mayor peso que los más específicos. Esta medida de similitud viene dada por el <i>CI</i> del pariente común más cercano (<i>PCC</i>) ²¹ .
$dist_{jc}(c_1, c_2) = CI(c_1) + CI(c_2) - 2 * CI(pcc(c_1, c_2))$	Esta medida de similitud propuesta por Jiang & Conrath [66], realiza una combinación entre los métodos basados en nodos y los métodos basados en enlaces. Además se basa en la medida planteada

²¹ En una jerarquía, representa a aquel concepto antecesor más cercano que es pariente de ambos conceptos.

	por Rada, con la diferencia de que este enfoque incorpora pesos en los nodos, para ser usados como factor de decisión.
$rel_{HS}(c_1, c_2) = C - len(c_1, c_2) - k * d$	Esta medida propuesta por Hirst [67], tiene en cuenta varios tipos de relaciones de una ontología tales como meronimia ²² , holonimia ²³ , hiponimia ²⁴ , hiperonimia ²⁵ , sinonimia ²⁶ y atonimia ²⁷ , pero se enfoca hacia una medida más general de la relación semántica. Además involucran constantes como C , k y d , las cuales sirven para ajustar la distancia de la ruta y los cambios de dirección de los enlaces.
$Sim_{tbk}(c_1, c_2) = \frac{2N}{N1 + N2} * PF(c_1, c_2)$	Esta medida propuesta por Slimani [68], está basada en el método de conteo de relaciones y consiste en una adaptación de la medida original de Wu y Palmer [69].

Tabla 11. Medidas de Similitud Semántica.

2.6 METABUSCADORES

Según los trabajos presentados en [70-71], existen múltiples definiciones en lo que a meta buscador se refiere, las más comunes son:

- Software que agrega los resultados de varios motores de búsqueda²⁸ o directorios web²⁹ para encontrar las páginas más relevantes.
- Elemento software que realiza sus búsquedas sobre auténticos buscadores, para luego analizar, procesar y presentar sus propios resultados.

Para el presente proyecto, una definición más acertada y por ende más simplificada, sería dar por entendido que un metabuscador es simplemente un buscador que funciona sobre otro buscador o buscadores.

²² Merónimo se denomina a la palabra cuyo significado constituye una parte del significado total de otra palabra.

²³ Holónimia es una noción semántica que se opone a meronimia, del mismo modo en que se oponen el todo y la parte.

²⁴ Se denomina hipónimo a aquella palabra que posee todos los rasgos semánticos, de otra más general.

²⁵ se denomina hiperónimo a aquel término general que puede ser utilizado para referirse a la realidad nombrada por un término más particular.

²⁶ Sinonimia es una relación de semejanza de significados entre determinadas palabras.

²⁷ La antonimia es un tipo de relación semántica que se establece entre palabras que poseen significados totalmente contrarios.

²⁸ Motor de Búsqueda: Un motor de búsqueda es un software que busca sitios web basándose para ello, en palabras clave designadas como términos de búsqueda

²⁹ Directorio Web: Un directorio web es un tipo de sitio web, que contiene un directorio organizado de enlaces a otros sitios web, con una estructura de categorías y subcategorías.

Características

- No poseen base de datos propia.
- Permiten realizar búsquedas en varios buscadores al mismo tiempo.

Ventajas

- Al realizar búsquedas sobre distintas fuentes de datos (buscadores), los meta buscadores, nos permiten ver la verdadera relevancia de una página web frente a otras.

Desventajas

- Ya que realizan múltiples consultas sobre distintos buscadores, los resultados suelen tardar un poco más que los de un buscador tradicional.

Teniendo en cuenta lo anterior, podría decirse que la arquitectura básica, de un meta buscador, es la mostrada en la Figura 4:

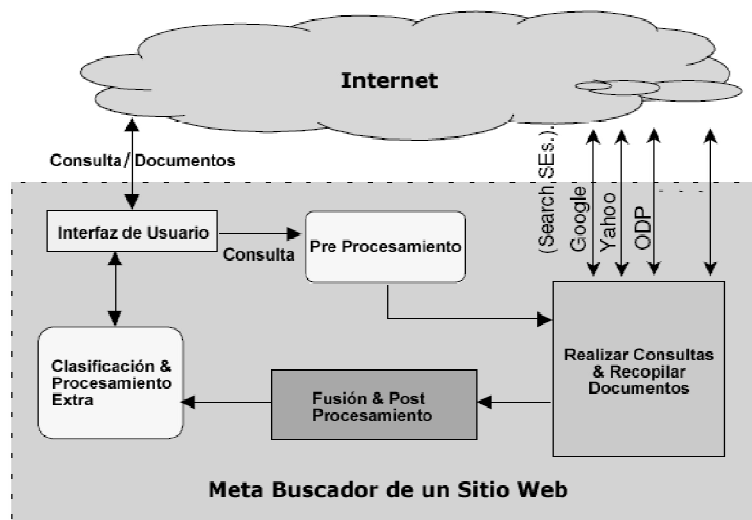


Figura 4. Arquitectura Básica de un Meta buscador.

Fuente: M. Manoj [71]

De acuerdo a la Figura 4, podría describirse en términos generales, el funcionamiento de un metabuscador, como la ejecución en forma secuencial, de las siguientes tareas:

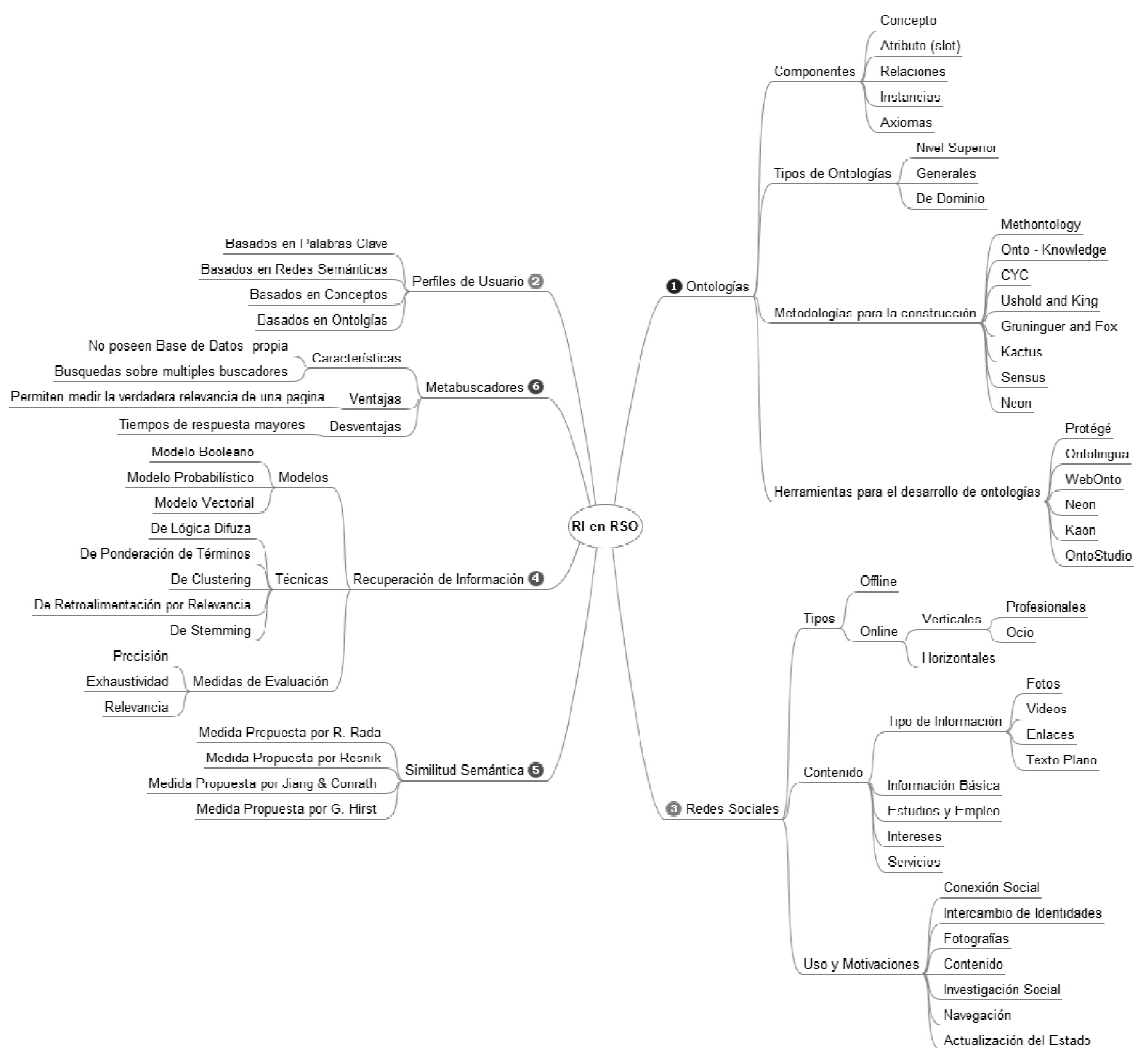
1. Recibir la consulta del usuario.
2. Procesar la consulta.
3. Realizar múltiples consultas.
4. Recoger y combinar los resultados.
5. Presentar los resultados pre procesados al usuario.

Algunos de los meta buscadores más populares son:

- DogPile³⁰
- Mamma³¹
- Vivisimo³²
- Kartoo³³
- Meta Crawler³⁴

2.7 SINTESIS MARCO TEORICO

La Figura 5 muestra gráficamente, los temas o ítems estudiados en este capítulo, con el objetivo de que el lector identifique los elementos que serán importantes para la posterior definición del modelo.



³⁰ <http://www.dogpile.com/>

³¹ <http://www.mamma.com>

³² <http://vivisimo.com/>

³³ <http://www.kartoo.com>

³⁴ <http://www.metacrawler.com/>

Figura 5. Elementos Principales para la Definición del Modelo

Los elementos presentados en la síntesis anterior, pueden ser agrupados en tres áreas temáticas generales, como son: Redes Sociales Online, Perfil de Usuario y Ontologías y la Recuperación de Información. Ahora bien, los trabajos relacionados estudiados anteriormente, han trabajado sobre estas 3 áreas temáticas, sin embargo en la literatura consultada no se encontró un estudio, que integrara de manera sinérgica el Perfil de Usuario basado en Ontologías y la Recuperación de Información aplicado a los sitios de Redes Sociales Online, esto representa un vacío investigativo, tal como lo muestra la zona sombreada en la Figura 6, y que puede ser abordado por el presente proyecto.



Figura 6. Trabajos Relacionados

2.8 CONCLUSIONES

- En la literatura consultada, no se encontró un estudio que abordara el tema de la RI en las RSO, más específicamente, que involucrara el perfil de usuario y las ontologías. Por esta razón, se observa que el planteamiento de un modelo, que integre perfil de usuario y ontologías, a fin de mejorar la relevancia de la información recuperada en las RSO, sería de gran utilidad.
- De acuerdo al estudio realizado en el presente capítulo, se puede identificar, que la ontología más adecuada para el presente proyecto, es la propuesta por Golemati et al [38], debido a las siguientes razones:
 - Está disponible su versión en inglés, para ser utilizada por el público en general.
 - Esta ontología es genérica y puede ser utilizada para representar cualquier tipo de perfil de usuario, ya que una característica importante, es que puede ser extensible y adaptable de acuerdo a las necesidades.

- Teniendo en cuenta el estudio previo, mostrado en el capítulo 2, se pudo observar que esta ontología, cubre la mayoría de los atributos que conforman el perfil de usuario de las RSO, que han sido fuente de nuestro estudio.
- Se utilizara una ontología para representar el perfil de usuario, debido a las siguientes razones:
 - Son útiles para estructurar conocimiento sobre determinada área. Esto nos permitirá, definir y estructurar de forma jerárquica, todos aquellos elementos que integraran el perfil de usuario.
 - Dado que uno de los puntos claves del proyecto, es el concerniente a la actualización del Perfil de Usuario, en base a la interacción que éste tenga con la RSO. El uso de una ontología cobra una importancia significativa, ya que al identificar el recurso (de la RSO sobre el cual el usuario ha mostrado interés) dentro de la estructura ontológica, es posible realizar inferencia hacia arriba (ascendente) o hacia abajo (descendente), con el objetivo de identificar que otros conceptos similares pueden ser activados (conceptos súper clase y/o sub clase respectivamente). Esto en algunos estudios es conocido como mecanismo de activación por propagación, y ha sido utilizado para lograr la correcta actualización del perfil de Usuario en base a la interacción con determinado sistema.
- La importancia de utilizar la similitud semántica en el proyecto, radica en que, gracias a esta medida es posible identificar conceptos que guardan cierta similitud y propagar sus pesos entre sí, lo cual sería de gran utilidad en el proceso de actualización de un PU, que utilice conceptos ponderados para representar el grado de interés del usuario en determinado tema.
- Se optara por utilizar la medida de similitud semántica propuesta por Slimani, a razón de su simplicidad de aplicación y rendimiento.

3 DEFINICION DEL MODELO BUSQUEDA DE INFORMACION EN REDES SOCIALES EN LINEA – MOBIRSE

El presente capítulo muestra la propuesta denominada MOBIRSE, un modelo de recuperación de información en Redes Sociales en Línea, el cual está enfocado en el uso de perfil de usuario y ontologías. En este capítulo se presenta la definición del modelo, el cual incluye: elementos, restricciones y módulos. Además se muestran las sugerencias para su aplicación, cabe resaltar que la metodología utilizada para generar el modelo, fue la propuesta por [72].

Para llevar a cabo la creación del modelo es necesario tener en cuenta 3 fases:

- **Fase de Conceptualización:** en la cual se determina el ámbito del modelo.
- **Fase de Formulación:** en la cual se define el modelo.
- **Fase de Evaluación:** en la cual se diseña la prueba preliminar del modelo.

3.1 FASE DE CONCEPTUALIZACIÓN

Esta fase es muy importante, puesto que permite identificar los elementos, actores y relaciones entre estos, que serán tenidos en cuenta a lo largo de la definición del modelo. A continuación se muestran los elementos y restricciones fundamentales que harán parte del modelo propuesto.

3.1.1 Elementos Fundamentales Del Modelo

- **Buscador RSO:** representa el servicio de búsqueda proporcionado por la RSO.
- **Recursos:** representa aquellos elementos que están almacenados en la RSO, tales como: fotos, videos, enlaces, texto plano.
- **API RSO:** representa el elemento mediante el cual, se accede a los recursos y servicios de la RSO.
- **Perfil de Usuario:** representa los gustos y preferencias del usuario a través de una ontología, la cual se define el formato estándar OWL [73].
- **Mecanismo RI:** representa una serie de procesos, que se le aplican a los recursos devueltos por el buscador de la RSO, con el fin de mejorar la relevancia de éstos.
- **Consulta:** representa la consulta original ingresada por el usuario.
- **Mecanismo de Expansión de Consulta:** representa una serie de procesos, que se aplican a la consulta original, a fin de complementarla para obtener resultados de búsqueda más relevantes para el usuario.

3.1.2 Restricciones Del Modelo

De acuerdo al estudio realizado en el capítulo anterior, se observa que la definición del modelo, estaría sujeta a las siguientes limitaciones:

- En principio, está orientado únicamente a los sitios de Redes Sociales Online (RSO).
- La RSO debe proveer una API de desarrollo, que tenga acceso tanto a la información del perfil de usuario, recursos de actualización y resultados de búsqueda, puesto que es esta información la que será procesada por el modelo.
- Los recursos que se tendrán en cuenta para, la actualización del perfil de usuario, serán aquellos que contengan texto plano como publicaciones, eventos, descripciones de páginas, etc.
- La información proveniente de la RSO debe estar en el idioma inglés, puesto que para su procesamiento, se hace necesario utilizar el diccionario de términos de WordNet³⁵, el cual es un recurso léxico que se encuentra en el idioma inglés.

3.2 MODELO CONCEPTUAL

La importancia de los modelos conceptuales, radica en abstraer los elementos relevantes de una realidad específica, a fin de ser analizados y comprendidos de una manera más sencilla. Por lo tanto, un modelo conceptual puede definirse, como una representación simplificada de sistemas o elementos del mundo real, que facilita la comprensión y entendimiento de éstos [74].

De acuerdo con lo anterior en la Figura 7, se muestran los módulos que harán parte de la definición de MOBIRSE y las relaciones entre éstos, generando así una vista global, de la primera aproximación al mismo.

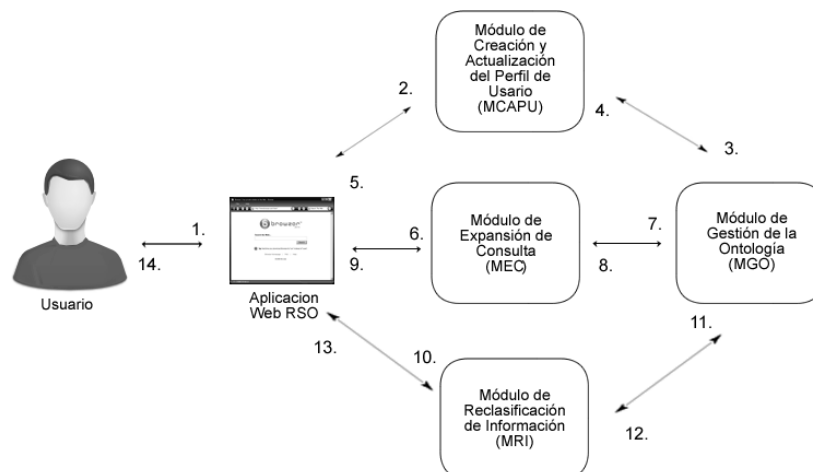


Figura 7. Vista global del Modelo RI en RSO

³⁵ <http://wordnet.princeton.edu/>

- **Aplicación Web RSO (App Web):** Es la aplicación sobre la RSO específica (facebook, twitter, etc.), que se encarga de capturar las acciones solicitadas por el usuario e interactuar con el modelo.
- **Modulo de Creación y Actualización del Perfil de Usuario (MCAPU):** Es el encargado de crear y actualizar la ontología que representa el perfil de usuario.
- **Modulo de Expansión de Consulta (MEC):** Es el encargado de procesar la consulta original, proveniente de la App Web y retornar los conceptos expandidos alrededor de la consulta dada.
- **Modulo de Gestión de la Ontología (MGO):** Es el encargado de acceder, salvar, inferir y demás operaciones que se puedan realizar sobre la ontología.
- **Modulo de Reclasificación de Información (MRI):** Es el encargado de ordenar por relevancia con base al Perfil de Usuario (PU), los resultados arrojados por el buscador de la RSO.

A continuación, se explica la interacción entre cada uno de esos módulos (según la Figura 7):

Paso 1. El usuario interactúa (registra, accede, realiza búsqueda) con la App Web.

Paso 2. La App Web se comunica con el Modulo de Creación y Actualización del Perfil de Usuario (MCAPU), teniendo en cuenta los siguientes casos:

Caso 1. El Usuario se registra en la App Web.

La App Web envía la información del PU y de interacción con la RSO (publicaciones, etiquetado, etc.), con el objetivo de que el Perfil de Usuario Ontológico sea creado y actualizado respectivamente.

Paso 3. El MCAPU solicita una instancia en blanco o vacía de la ontología, al Modulo de Gestión de la Ontología (MGO), a fin de realizar las operaciones descritas anteriormente.

Paso 4. El MGO envía la instancia en blanco solicitada, para crear y actualizar el perfil de usuario.

Caso 2. El Usuario ingresa a la App Web.

La App Web envía la información de interacción con la RSO (publicaciones, etiquetado, etc.), con el objetivo de que el perfil de usuario ontológico pueda ser actualizado.

Paso 3. El MCAPU solicita al MGO, una instancia de la ontología cargada con la información del usuario, que se encuentra actualmente utilizando la Aplicación Web de la RSO.

Paso 4. El MGO carga la información del PU desde la base de datos en una instancia de la ontología, a fin de que esta pueda ser actualizada con los recursos enviados previamente por la App Web.

Caso 3. El usuario realiza una consulta.

La App Web envía la consulta MCAPU, a fin de utilizarla para actualizar el perfil de usuario ontológico.

Paso 3. El MCAPU envía la consulta proveniente de la App Web al MGO, para actualizar el perfil de usuario.

Paso 4. El MGO realiza la respectiva actualización del perfil de usuario.

Paso 5. El Modulo de Creación y Actualización del Perfil de Usuario envía a la App Web, la información solicitada teniendo en cuenta los casos anteriores.

Paso 6. La App Web envía la consulta al Modulo de Expansión de Consulta (MEC).

Paso 7. El MEC solicita la ontología del PU al MGO, con el fin de utilizar el perfil de usuario en el proceso de expansión de la consulta.

Paso 8. El MGO retorna la ontología solicitada por el MEC.

Paso 9. El MEC identifica los conceptos más similares a la consulta, y los envía a la App Web, para que esta última los utilice para refinar la consulta original.

Paso 10. Teniendo en cuenta, la consulta expandida por el MEC, la App Web, obtiene los resultados de búsqueda de la RSO y los envía al MRI.

Paso 11. El MRI solicita la ontología del PU al MGO, con el objetivo de reclasificar los documentos con base al perfil de usuario ontológico.

Paso 12. El MGO retorna la ontología solicitada por el MRI.

Paso 13. Se reclasifican los documentos recuperados, teniendo en cuenta los pesos de los conceptos presentes en el perfil de usuario ontológico y se envían a la App Web. Para ello se aplica el algoritmo propuesto por Sieg et al [8], el cual es una variación del modelo vectorial.

Paso 14. La App Web despliega al usuario los documentos reclasificados.

3.3 FASE DE FORMULACION

En esta fase los elementos identificados anteriormente son representados mediante un lenguaje formal, para los cuales se hace necesario diagramas y ecuaciones que serán manejadas en el modelo [75].

A continuación se describe en detalle el Modelo de Búsqueda y Recuperación de Información en Redes Sociales en Línea (MOBIRSE).

3.3.1 Aplicación Web RSO (App Web)

Es la aplicación encargada de capturar las acciones solicitadas por el usuario y desplegar la información proveniente de los otros módulos. Además, es la que hace uso de la API de la RSO específica, para obtener la información del PU, de interacción con la RSO y los resultados de búsqueda. La Figura 8 muestra la interacción de la App Web con los módulos de MOBIRSE.



Figura 8. Aplicación Web RSO

- La App Web debe comunicarse con el MCAPU para:
 - Cargar la información del perfil de usuario de la RSO en la ontología, creando así, el perfil de usuario ontológico, para un usuario en particular.
 - Actualizar el perfil de usuario ontológico, de acuerdo a la información generada por la interacción del usuario con la RSO.
- La App Web debe comunicarse con el MEC para:
 - Extraer términos del PU que guardan cierta similitud con la consulta.
- La App Web debe comunicarse con el MRI para:
 - Ponderar los resultados obtenidos a través del buscador de la RSO.

Cabe aclarar, que la información de la RSO obtenida por la App Web, es la que alimenta el modelo y debe ser utilizada, según se explico anteriormente.

3.3.2 Modulo de Construcción y Actualización del Perfil de Usuario (MCAPU)

Este modulo es el encargado de crear y actualizar la ontología que representa el perfil de usuario. Para llevar a cabo lo anterior, se hará uso de la ontología propuesta por Golemati et al [38], la cual en su versión original se compone de los conceptos mostrados en la

Tabla 12, la cual fue adaptada y ampliada, teniendo en cuenta las necesidades de nuestra investigación.

Concepto	Descripción
Person (<i>user</i>)	Representa la información básica del usuario, como nombre, apellido, fecha de nacimiento, etc.
Characteristic	Representa características generales del usuario, tales como, color de ojos, peso, altura, etc.
Ability	Representa habilidades y discapacidades, tanto físicas y mentales del usuario.
Living Conditions	Representa el lugar de residencia y las condiciones de vida del usuario.
Contact (<i>relation</i>)	Representa las relaciones del usuario con otras personas y el tipo de estas relaciones.
Preference	Representa las preferencias del usuario, como por ejemplo, "amor a los gatos", "le gusta el color azul", etc.
Interest	Representa los intereses de la persona, por ejemplo, "interés en deportes", "interés en cocina", etc.
Activity (<i>Hobby, Research/Interest</i>)	Representa hobbies o actividades relacionadas con el trabajo, por ejemplo, "coleccionar estampitas", "investigar las cruzadas", etc.
Education	Representa los niveles de educación del usuario, por ejemplo un título universitario o un título en lenguas.
Profession	Representa las profesiones del usuario.
Expertise	Representa la destreza del usuario una actividad determinada, por ejemplo, experto en computación.
Thing	Representa las cosas vivas y no vivas, que el usuario posee o no y que están relacionadas con el, por ejemplo un carro o una mascota.

Tabla 12. Conceptos de la Ontología del Perfil de Usuario

Cabe aclarar que los cambios realizados sobre la ontología fueron los siguientes:

- Se agregaron al concepto *Person* los atributos: *id*, *nickname*, *biography*, e igualmente se agrego una referencia al concepto *Characteristics*, puesto que a raíz del estudio realizado en el Capítulo 1, se identifico que dichos atributos hacen parte de la mayoría de los perfiles de usuario de las RSO.
- Se agrego el atributo *score* a todos los conceptos de la ontología, puesto que cada concepto de la ontología necesita tener un peso que representa la importancia para el usuario.

A continuación, en la Figura 9 se muestra el funcionamiento general de este modulo:

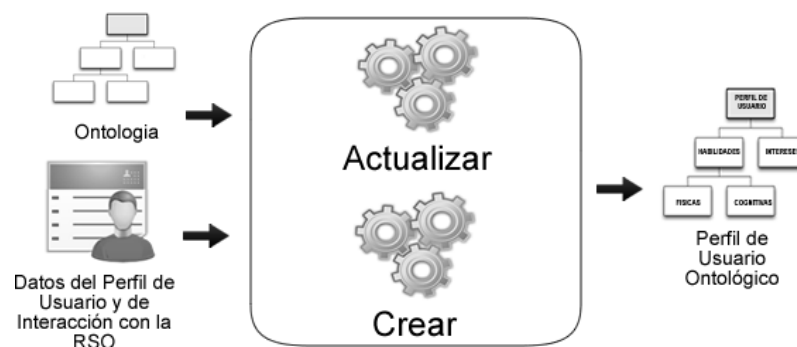


Figura 9. Módulo de Construcción y Actualización del Perfil de Usuario

3.3.2.1 Creación del Perfil de Usuario

La creación del PU para cada usuario se realizara la primera vez que la App Web interactué con el modelo, para el usuario en cuestión, puesto que para posteriores interacciones, la información del PU será cargada desde un repositorio de base de datos.

Teniendo en cuenta lo anterior, para la creación del perfil de usuario, se realizan las siguientes acciones:

1. La App Web, envía los datos del PU de la RSO al MCAPU. Estos datos son extraídos por medio de la API de la RSO.
2. El MCAPU, solicita una nueva instancia de la ontología al MGO, la cual contiene todos los pesos de los conceptos, con un valor inicial en 0.01, dicho valor de inicialización fue seleccionado, ya que no podía ser cero, puesto que si al utilizar por primer vez el modelo, no existiesen recursos de actualización, el modulo de reclasificación asignaría a los documentos un valor de 0. No se fijo un valor de 1, puesto que dicho valor representa el puntaje máximo que pueden alcanzar los pesos de los conceptos de la ontología, lo cual significaría que inicialmente todos los conceptos de la ontología son de interés para el usuario, lo cual es erróneo puesto que el interés que estos representan para el usuario se ajustan a medida que el usuario interactúa con la red social.
3. El MCAPU, recorre los datos del PU y los carga en la instancia de la ontología.

3.3.2.2 Actualización del Perfil de Usuario

Para la actualización del perfil de usuario, se hace necesario definir los servicios de la RSO, que serán utilizados como fuente de actualización del PU. Según el estudio realizado en el capítulo anterior, se observa que servicios como publicaciones, eventos, etc. (para mayor información, ver la Tabla 9) pueden ser útiles en el momento de actualizar el PU.

Para actualizar el PU, se realizan los siguientes pasos:

1. Si existe información de interacción del usuario con la RSO (publicaciones, búsquedas, etc.), extraída por medio de la API de la RSO, la App Web, envía una

colección de términos ponderados al MCAPU, extraídos mediante lógica propia de la aplicación o haciendo uso de un servicio del modelo, el cual se basa en la propuesta hecha por Baziz et al [76], a fin de identificar conceptos y que tiene como entrada una colección de texto plano. Cabe aclarar, que dicha colección de términos ponderados, representaría la más reciente interacción del usuario con la RSO.

2. Teniendo en cuenta la colección de términos ponderados del paso 1, se procede a actualizar los pesos de los conceptos de la ontología. Para llevar a cabo esta operación, se hace uso del algoritmo de activación por propagación, propuesto por Alexander Trousov et al [77], el cual permite identificar y activar conceptos que guardan cierto grado similitud, a partir de un conjunto inicial de conceptos identificados en la ontología. Dicha similitud será cuantificada, a través de la medida de similitud semántica definida por Slimani et al [68]. El proceso para la actualización del PU, se muestra a continuación:

1. El PU es cargado a través del MGO.
2. Se identifican los conceptos que se emparejan con los términos de la colección enviada por la App Web, teniendo en cuenta sinónimos provistos por WordNet. Los términos que no se emparejen con ningún concepto de la ontología, son procesados para buscar algún tipo de relación jerárquica con los conceptos existentes, utilizando hipónimos e hiperónimos. De existir dicha relación, se agrega el nuevo nodo y se actualiza la estructura ontológica, de no existir, pasaran a ser agregados como nuevos nodos hijos del concepto "*Others*", el cual ha sido definido para albergar conceptos, que representan intereses del usuario, pero que no tienen relación alguna con los conceptos existentes.
3. Se asignan los pesos de la colección a los conceptos que fueron procesados en el paso anterior.
4. Los conceptos procesados son enviados al MGO, el cual mediante el mecanismo de activación por propagación, se encarga de propagar pesos a conceptos de la ontología, que guardan cierta similitud con los conceptos identificados.

3.3.3 Modulo de Gestión de la Ontología (MGO).

Este modulo es el encargado de realizar toda la gestión con respecto a la ontología, entendiéndose por gestión, a las funciones para acceder, salvar, inferir y demás operaciones que se puedan realizar sobre ésta. En la Figura 10 se muestra el funcionamiento general de este módulo:

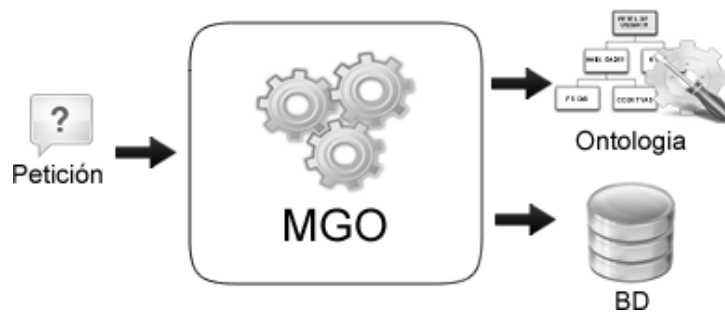


Figura 10. Módulo de Gestión de la Ontología.

Las funciones que realiza este modulo, varían dependiendo del tipo de petición que se le solicite. Las funcionalidades que éste ofrece son:

- **Obtener una nueva instancia de la ontología**

Retorna una instancia de la ontología en blanco, para crear un nuevo PU.

- **Cargar el PU desde la base de datos**

Carga la información del PU a una instancia de la ontología, desde un repositorio de base de datos, que será utilizado para salvar y recuperar el estado del perfil de usuario ontológico. Cabe aclarar que este esquema de base de datos relacional, es una réplica exacta de la ontología que representa el Perfil de Usuario.

- **Salvar información del PU**

Mapea la información del Perfil de Usuario Ontológico, al esquema de base de datos relacional, que fue descrito en la funcionalidad anterior.

- **Propagar Pesos**

Es la funcionalidad más importante de este modulo, y se encarga de actualizar los pesos de los conceptos de la ontología, haciendo uso de la medida de similitud semántica propuesta por Slimani y la técnica de activación por propagación, propuesta por Trussov.

Dado que se debe determinar qué porcentaje de peso del concepto identificado debe propagarse a conceptos similares, se propone la siguiente medida, la cual está basada en la medida de similitud semántica y el peso que contiene el concepto al cual se le realiza el proceso de propagación.

Teniendo en cuenta lo anterior, el peso propagado del concepto C_i al concepto C_j , está dado por la siguiente expresión:

$$\text{PesoPropagado}(C_i, C_j) = \text{Similitud}(C_i, C_j) * \text{Peso}(C_i) \quad \text{Ecuación 1.}$$

Donde:

C_i Representa el concepto identificado en la ontología, al que se le quiere aplicar el proceso de propagación.

C_j Representa, un concepto cualquiera de la ontología al que se le pretende propagar un porcentaje del peso de C_i .

$Similitud(C_i, C_j)$ Representa el valor de similitud semántica del concepto C_i con respecto al concepto C_j .

$Peso(C_i)$ Representa el peso o puntaje del concepto C_i

De la Ecuación 1, vale la pena aclarar que un mayor valor de similitud entre los dos conceptos, aporta un mayor porcentaje de peso propagado y viceversa. Además, dado que tanto el valor de similitud como el peso están acotados en un rango de 0 a 1, el valor del peso propagado estará también en dicho rango.

Obtenido el valor de propagación con base en la Ecuación 1, se debe actualizar o reasignar el peso que tendrá el concepto C_j , lo cual se logra calculando el promedio de la suma del peso actual de C_j con el peso propagado por C_i . Teniendo en cuenta lo anterior, el nuevo peso de C_j , estará dado por la siguiente expresión:

$$Peso(C_j) = (PesoPropagado(C_i, C_j) + Peso(C_j)) / 2 \quad \text{Ecuación 2}$$

En la Ecuación 2, se puede observar que se realiza un promedio entre el peso actual que tiene el concepto C_j y el peso que le ha sido propagado. Esta operación fue necesaria, ya que se debería tener en cuenta, el peso histórico que tenía el concepto al momento de intentar actualizar su valor.

Finalmente cabe aclarar que si el valor de propagación es igual a cero, no se aplicara la Ecuación 2 y el concepto conservara su peso actual.

Aclarada la forma como se reasignan o actualizan los pesos de los conceptos de la ontología, el proceso general de propagación de pesos sobre la ontología, se detalla a continuación:

Para cada concepto identificado:

1. Calcular el peso que este concepto propaga, a cada uno de los conceptos de la ontología, haciendo uso de la Ecuación 1.
2. Se extrae el mayor valor de propagación para cada concepto de la ontología, que haya sido afectado por los conceptos identificados.

3. Teniendo en cuenta los pesos obtenidos para los conceptos identificados en el paso 2, se procede a actualizar el peso de dichos conceptos, teniendo en cuenta la Ecuación 2.

A continuación se presenta un ejemplo que mostrara con mayor claridad el proceso descrito anteriormente.

Sean:

- $C_1 = \textit{music}$ con un peso de **0.6**
- $C_2 = \textit{musical genre}$ con un peso de **0.8**
- $C_3 = \textit{rock}$ con un peso de **0.9**
- $C_4 = \textit{pop}$ con un peso de **0.01**
- $C_5 = \textit{tv}$ con un peso de **0.01**
- $C_6 = \textit{tv show}$ con un peso de **0.01**
- $C_7 = \textit{friends}$ con un peso de **0.01**
- $C_8 = \textit{futurama}$ con un peso de **0.01**

Conceptos de la ontología que representa el perfil de usuario, tal como puede observarse en la Figura 11 Parte (A).

Ahora bien, se tiene como entrada el término **musical genre**, con un peso de 1. Dado que el término **musical genre**, está presente en la ontología y corresponde al concepto C_2 , se procede a propagar su peso a los conceptos más similares, para ello se hace necesario calcular la similitud entre C_2 y los demás conceptos de la ontología.

Haciendo uso de lo propuesto por Slimani, se procede a calcular la similitud semántica entre el concepto C_2 y los demás conceptos de la ontología, como se muestra a continuación:

$$\begin{aligned}
 Sim(C_2, C_1) &= 0.66 \\
 Sim(C_2, C_3) &= 0.8 \\
 Sim(C_2, C_4) &= 0.8 \\
 Sim(C_2, C_5) &= 0.0 \\
 Sim(C_2, C_6) &= 0.0 \\
 Sim(C_2, C_7) &= 0.0 \\
 Sim(C_2, C_8) &= 0.0
 \end{aligned}$$

Teniendo en cuenta las similitudes calculadas anteriormente y la Ecuación 1, se procede a calcular el peso que C_2 propaga a los demás conceptos. Estos valores se muestran a continuación:

$$\begin{aligned}
 PesoPropagado(C_2, C_1) &= 0.66 \\
 PesoPropagado(C_2, C_3) &= 0.8 \\
 PesoPropagado(C_2, C_4) &= 0.8 \\
 PesoPropagado(C_2, C_5) &= 0.0 \\
 PesoPropagado(C_2, C_6) &= 0.0
 \end{aligned}$$

$$\begin{aligned} \text{PesoPropagado}(C_2, C_7) &= 0.0 \\ \text{PesoPropagado}(C_2, C_8) &= 0.0 \end{aligned}$$

Ahora bien, teniendo en cuenta estos valores de propagación y la Ecuación 2, se procede a actualizar el peso de los demás conceptos, lo cual se muestra a continuación:

$$\begin{aligned} \text{Peso}(C_1) &= 0.6 \\ \text{Peso}(C_2) &= 1 \\ \text{Peso}(C_3) &= 0.85 \\ \text{Peso}(C_4) &= 0.405 \\ \text{Peso}(C_5) &= 0.01 \\ \text{Peso}(C_6) &= 0.01 \\ \text{Peso}(C_7) &= 0.01 \\ \text{Peso}(C_8) &= 0.01 \end{aligned}$$

Teniendo en cuenta los valores calculados anteriormente, la Figura 11 parte (B), muestra el nuevo estado del perfil de usuario, tras el proceso de propagación de pesos.

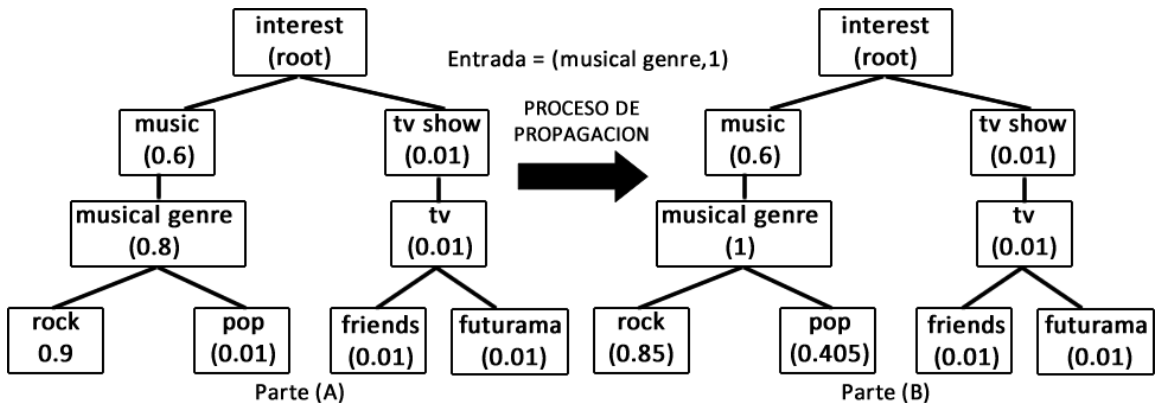


Figura 11. Ejemplo de Propagación de Pesos.

- Extraer Términos

Es la función encargada de extraer términos simples y compuestos contenidos en una fuente de texto plano como: publicaciones, eventos, mensajes, consultas, etc. Es necesario precisar, que la forma como opera esta funcionalidad, se basa en lo propuesto por Baziz [76].

Para dar mayor claridad, se explicará mediante un ejemplo, la forma como la propuesta de Baziz, hace procesamiento de texto de izquierda a derecha, a fin de extraer términos simples y compuestos presentes en una entrada de texto. Para ello, el proceso consta de lo siguiente:

Por cada iteración se van suprimiendo términos de izquierda a derecha del texto original hasta que la cadena se componga de un solo término o palabra. Seguidamente se eliminara dicho termino de la entrada original, y se inicia la siguiente

iteracion. El proceso termina una vez que la entrada se compone de un solo término o palabra.

A continuación se muestra a modo de ejemplo, su funcionamiento, mostrando el estado de la entrada de texto por cada iteración:

Consideremos, la siguiente entrada de texto: **music video by Papa Roach**

<p>Iteracion 1 <i>music video by Papa Roach</i> <i>music video by Papa</i> <i>music video by</i> <i>music video</i> <i>music</i></p>	<p>Iteracion 2 <i>video by papa roach</i> <i>video by papa</i> <i>video by</i> <i>video</i></p>
<p>El termino music será suprimido para la iteracion 2</p>	<p>El termino video será suprimido para la iteracion 3</p>
<p>Iteracion 3 <i>by papa roach</i> <i>by papa</i> <i>by</i></p>	<p>Iteracion 4 <i>papa roach</i> <i>papa</i></p>
<p>El termino by será suprimido para la iteracion 4</p>	<p>El termino papa será suprimido para la iteracion 5</p>
<p>Iteracion 5 <i>roach</i></p>	
<p>Fin del proceso puesto que la entrada de texto se compone de un solo termino</p>	

De lo anterior, puede observarse que términos como: **music**, **video**, **music video**, **papa roach**, podrían representar algún interés o preferencia para determinado usuario.

Teniendo en cuenta lo anteriormente descrito, y el hecho de que se tendrán en cuenta aquellos términos que no sean Palabras Vacías³⁶ ni verbos. El proceso para extraer términos se describe a continuación:

1. Se extraen cada uno de los términos junto con su frecuencia de aparición en el documento, lo cual produce una colección de términos **T** como la siguiente:

$$T = \{(t_1, f_1), (t_2, f_2), \dots, (t_n, f_n)\}$$

Donde:

t_i Representa el término simple o compuesto extraído,

³⁶ Palabras Vacías: es el nombre que reciben las palabras sin significado como artículos, pronombres, preposiciones, etc. En el idioma inglés estas palabras son conocidas como stop words.

$$\forall_i, i = 1, \dots, n$$

f_i Representa la frecuencia del término t_i , $\forall_i, i = 1, \dots, n$

2. Para cada término t_i identificado en el paso anterior, se normaliza su frecuencia f_i , mediante la Normalización Min-Max [78], tal como lo muestra la siguiente expresión:

$$p_i = \frac{f_i - \min}{\max - \min} \quad \text{Ecuación 3}$$

Donde:

f_i Representa el valor de la frecuencia del término t_i , $\forall_i, i = 1, \dots, n$

\max Representa el mayor valor de frecuencia de la colección de términos.

\min Representa el menor valor de frecuencia de la colección de términos.

Lo anterior es necesario, ya que permitirá tener términos con pesos en un rango de 0 a 1.

Como resultado del proceso anterior, se obtendrá una colección de términos ponderados TP , tal como se muestra a continuación:

$$TP = \{(t_1, p_1), (t_2, p_2), \dots, (t_n, p_n)\}$$

Donde:

t_i Representa el término simple o compuesto extraído, $\forall_i, i = 1, \dots, n$

p_i Representa el peso normalizado del término t_i , $\forall_i, i = 1, \dots, n$

Finalmente, cabe resaltar, que esta colección de términos extraída, será la utilizada para propagar pesos en la ontología.

- Agregar Concepto

Función encargada de agregar nuevos conceptos a la estructura ontológica, para esto se hará uso de términos hiperónimos³⁷, hipónimos³⁸ y sinónimos³⁹ provistos por

³⁷ Términos Hiperónimos: son aquellos, que por tener un significado más amplio, incluyen otros más concretos o específicos, como por ejemplo la palabra *flor* es un hiperónimo.

³⁸ Términos Hipónimos: contrario a los hiperónimos, son las palabras de significado restringido con las que se puede concretar a otras de significado más amplio, como por ejemplo las palabras *clavel*, *rosa*, *jazmín*, son hipónimos de *flor*.

³⁹ Sinónimos: son palabras que tienen un significado similar o idéntico entre sí.

WordNet. Para agregar un nuevo concepto a la ontología se han identificado 4 casos base:

1. No existe relación alguna entre el nuevo concepto y los presentes en la ontología.

Para este caso simplemente, se creara un nuevo nodo (concepto), con los atributos del término ponderado, es decir nombre y peso, y se agregara como hijo de la sección "Others". El siguiente grafico ilustra mejor este caso:

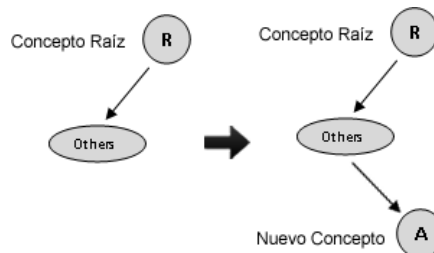


Figura 12. Agregar nuevo nodo, según el caso 1.

2. Existe un concepto hipónimo en la ontología y es hijo de un concepto raíz.

Se agrega el nuevo concepto como hijo del concepto raíz, y el nuevo concepto pasa a ser padre del concepto hipónimo.

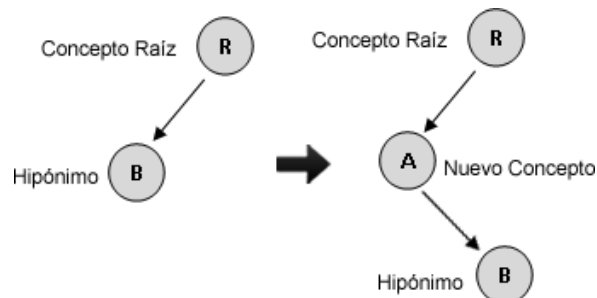


Figura 13. Agregar nuevo nodo, según el caso 2.

3. Existe en la ontología un concepto hiperónimo del nuevo concepto, que es nodo hoja.

El nuevo nodo es agregado como hijo del concepto hiperónimo.

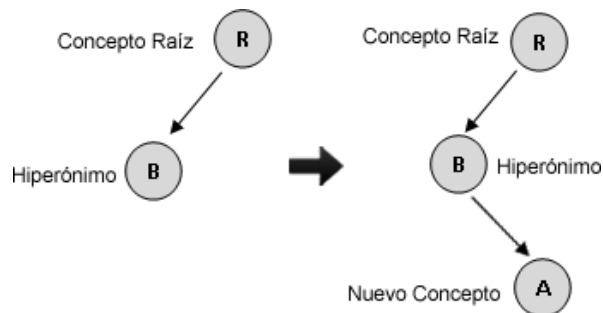


Figura 14. Agregar nuevo nodo, según el caso 2.

4. Existe en la ontología un concepto hipónimo e hiperónimo del nuevo concepto.

El nuevo concepto es agregado como nodo hijo del concepto hiperónimo, y el concepto hipónimo pasa a ser hijo del nuevo concepto.

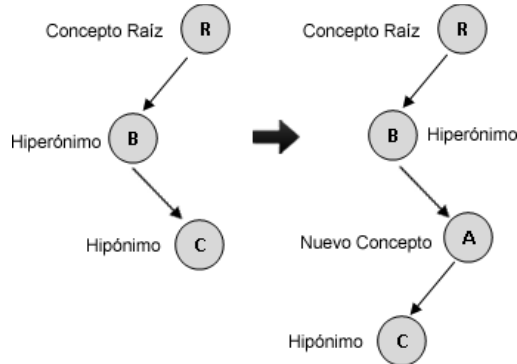


Figura 15. Agregar nuevo nodo, según el caso 4.

Finalmente, cabe aclarar que una vez que el nuevo nodo ha sido agregado a la ontología, se invoca a la función de propagar pesos, con el objetivo de que conceptos cercanos o similares queden actualizados.

3.3.4 Modulo de Expansión de Consulta (MEC).

Este modulo es el encargado de procesar la consulta original, proveniente de la App Web, para identificar conceptos del Perfil de Usuario Ontológico, que están relacionados con la consulta. Se hace necesario proveer este servicio, ya que ayudaría a mejorar o estructurar la consulta original y en consecuencia los resultados provenientes del buscador de la RSO. La Figura 16 muestra el funcionamiento general de este modulo.

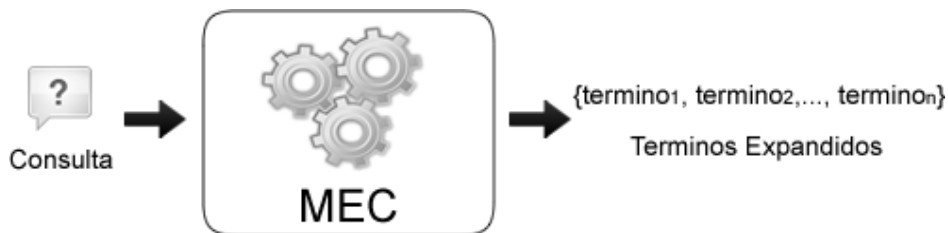


Figura 16. Modulo de Expansión de Consulta.

A fin de determinar los conceptos que serán devueltos a la App Web, se hará uso del modelo del *Espacio Vectorial*, para poder identificar el concepto mas parecido a la consulta, hecho esto, se procederá a extraer los hijos de dicho concepto a fin de devolver una lista de los conceptos más relacionados con la consulta original. A continuación se detalla el proceso a seguir:

Sean:

q	La consulta enviada desde la App Web.
$O = \{c_1, c_2, \dots, c_n\}$	El conjunto de conceptos que representan la ontología del Perfil de Usuario.
T	El vector de términos extraídos a partir de O .
$t_q = \{t_1, t_2, \dots, t_n\}$	El vector de términos que representa la consulta enviada desde la App Web en función de los términos contenidos en T .
$t_{ci} = \{t_1, t_2, \dots, t_n\}$	El vector de términos que representa cada concepto C_i , de la ontología, en función de los términos contenidos en T . $\forall_i, i = 1, \dots, n$
$Similitud(q, t_{ci})$	El valor de similitud dado como el coseno entre el vector de términos de la consulta q y el vector de términos del concepto C_i , $\forall_i, i = 1, \dots, n$
S	Array ⁴⁰ para almacenar tuplas de la forma $\langle Ci, Similitud(q, t_{ci}) \rangle$, con $Ci \in O, \forall_i, i = 1, \dots, n$
Te	Array que almacenara los términos expandidos que serán devueltos a la App Web.

A modo de pseudocódigo, a continuación se muestra el proceso general para obtener términos expandidos a partir de la consulta original.

```

Para cada  $C_i \in O$  haga
     $t_{ci} = crearVector(C_i, T)$ 
     $S.agregar(Ci, Similitud(t_q, t_{ci}))$ 
Fin Para
 $S.ordenar()$ 
 $Te = ExtraerHijos(S(0))$ 

```

De lo anterior, cabe aclarar que:

1. El array S , es ordenado de forma descendente a fin de extraer, la tupla con el mayor valor de similitud.
2. La función **ExtraerHijos**, se encarga de extraer los hijos del concepto de mayor valor de similitud con respecto a la consulta.
3. El array Te es enviado a la App Web.

⁴⁰ Array: representa de una colección de elementos.

3.3.5 Módulo de Reclasificación de Información (MRI)

Este módulo es el encargado de ponderar por relevancia los resultados de búsqueda, enviados por la App Web, los cuales son proporcionados por el servicio de búsqueda de la RSO. Para llevar a cabo esta operación, se debe tener en cuenta el PU del usuario en cuestión y la consulta que este ingresa. La Figura 17 muestra el funcionamiento general de este módulo.



Figura 17. Módulo de Reclasificación de Información.

Dado que se hace importante tener en cuenta el PU para reclasificar los documentos recuperados, se opta por utilizar una variante del modelo vectorial que utilice la ontología que representa el PU, para asignar pesos a los documentos, y por consiguiente reclasificarlos. Por tal motivo, el proceso de reclasificación de documentos se basará en lo propuesto por Sieg et al [8], el cual tiene en cuenta la evidencia semántica de la ontología, para asignar pesos a los documentos que se reclasificarán.

Para llevar a cabo este proceso es necesario aclarar:

- Que documentos, consulta y conceptos deben ser representados en función de vectores, que almacenan la frecuencia de aparición de los términos en los documentos recuperados, dicha frecuencia es calculada haciendo uso de la medida *tfidf* [55].
- Se utilizará la medida de similitud del coseno, para determinar el grado de similitud entre pares de vectores, dicho valor estará en un rango de 0 a 1.

Aclarado lo anterior, el proceso de reclasificación de documentos es como sigue:

Por cada documento recuperado, se realiza lo siguiente:

1. Se extrae el concepto de la ontología más similar al documento.
2. Se calcula la similitud del coseno:
 - Entre el documento y la consulta.

- Entre la consulta y el concepto más similar al documento.
3. Se asigna un puntaje al documento, teniendo en cuenta la similitud del documento con la consulta, la similitud de concepto más similar con la consulta y el peso de dicho concepto.

3.4 FASE DE EVALUACION

Teniendo en cuenta que a partir del modelo se desarrolla un prototipo software, la fase de evaluación del modelo está ligada a la implementación y posterior validación del prototipo (Capítulo 4 y Capítulo 7). Vale la pena aclarar, que el modelo presentado en la fase de formulación, es producto de una serie de iteraciones en su construcción, las cuales fueron soportadas por la validación que se le hizo al prototipo, y que permitió adicionar mejoras e igualmente eliminar elementos que no aportaban al modelo.

4 IMPLEMENTACION DEL MODELO

El desarrollo del prototipo se baso en el uso la metodología UP Ágil (Agile Unified Process⁴¹ [79]), utilizando sus fases y los artefactos que se consideren necesarios. Las fases son las siguientes:

4.1 FASE DE INICIO

En esta fase se definen los casos de uso más relevantes a un alto nivel con su respectivo diagrama de casos de uso, con el objetivo de establecer los requisitos mínimos para el desarrollo del prototipo de acuerdo a lo establecido en el modelo teórico. Posteriormente se define, a modo general una arquitectura potencial que permita guiar el proceso de desarrollo.

En general el proceso de desarrollo se llevo a cabo con la herramienta de desarrollo Visual Studio 2010 con Framework 3.5 de Microsoft .NET, utilizando el lenguaje de programación C# y motor de base de datos SQL Server Express Edition 2008.

Igualmente también se hizo necesario contar con los siguientes elementos:

- **Ontologías para representar el Perfil de Usuario:** cabe recordar que en la fase de definición del modelo, se opto por utilizar la ontología propuesta por Golemati.
- **Editores de Ontologías:** se hicieron pruebas con Protege 3.1.1 y Neón. Se opto por Protege, dado que el grupo de investigación tenía experiencia previa con el uso de esta herramienta y por el gran volumen de información y soporte disponible en la internet.
- **APIS para trabajar con ontologías desde el Framework de .NET:** se hizo un estudio previo con Jena .NET Framework [80] y OwlDotnetApi [81]. Se opto por utilizar Jena .NET Framework, dado que es una de las APIS que goza de mayor popularidad entre los investigadores de la web semántica, además porque dispone de gran cantidad de recursos de ayuda en la internet. Por último, cabe resaltar que experiencia previa en el grupo de investigación demuestra su utilidad y desempeño en el manejo de ontologías.
- **APIS para el procesamiento de Texto:** se opto por el API de Lucene para .NET [82], ya que es una de las más utilizadas en el área de la recuperación de la información y dispone de una serie de clases que facilitan entre otras: la partición de texto en palabras simples o términos y la remoción de palabras irrelevantes, vacías o stop words.

⁴¹ <http://cgi.una.ac.cr/AUP/html/overview.html>

- **Diccionarios de Términos:** se hizo uso del diccionario léxico WordNet a través de las API IKVM.OpenJDK.ClassLibrary⁴², JAWS⁴³ para la obtención de sinónimos, hiperónimos, hipónimos, etc.

Cabe aclarar que la implementación del modelo se hará a modo de librería, a fin de ser usado por cualquier aplicación web desarrollada en .NET.

4.1.1 Análisis de Requerimientos

Se desarrollará una librería, que pueda ser utilizada por una aplicación web de RSO, y que sea capaz de dar soporte a los siguientes servicios:

- Crear el perfil de usuario ontológico en base a la información del perfil de usuario de la RSO.
- Expandir la consulta proveniente de la aplicación web, haciendo uso de los intereses almacenados en el perfil de usuario ontológico.
- Reclasificar los resultados de búsqueda provenientes de la aplicación web, teniendo en cuenta el perfil de usuario ontológico.
- Actualizar el perfil de usuario ontológico en base a recursos de texto plano (publicaciones, eventos, consultas) que son proveídos por la RSO.
- Salvar o guardar el estado del perfil de usuario ontológico.

4.1.2 Diagrama de Casos de Uso

Los diagramas de casos de uso muestran el comportamiento de un sistema desde el punto de vista del usuario. Por tal motivo, son utilizados para determinar los requisitos funcionales del sistema, es decir, representan las funciones que un sistema puede ejecutar o realizar. La Figura 18, muestra el diagrama de casos de uso de MOBIRSE.

⁴² IKVM.OpenJDK.ClassLibrary: permite la ejecución de código Java sobre la plataforma .NET.

⁴³ JAWS: Api desarrollada en el lenguaje de Programación Java, que permite la recuperación de datos de la base de datos de WordNet.

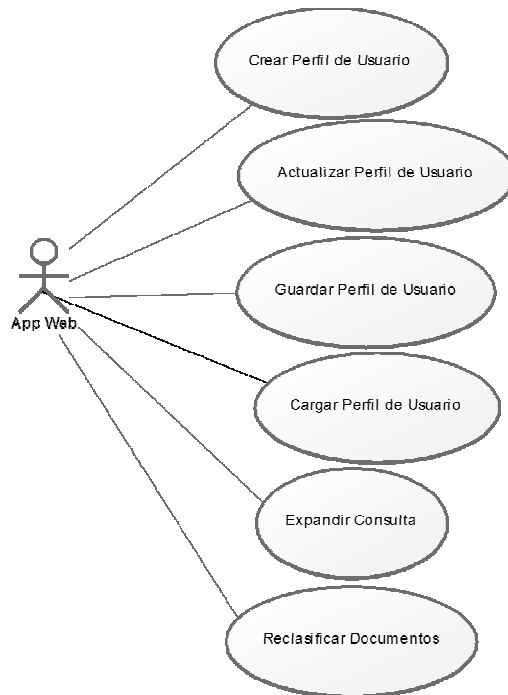


Figura 18. Diagrama de Casos de Uso

4.1.3 Casos de Uso en Formato Compacto

En ingeniería del software, un caso de uso es una técnica para la captura de requisitos potenciales de un nuevo sistema o una actualización de software. Cada caso de uso proporciona uno o más escenarios que indican cómo debería interactuar el sistema con el usuario o con otro sistema para conseguir un objetivo específico.

En la Tabla 13, se presentan los casos de uso en formato compacto.

CASO DE USO	Crear Perfil de Usuario
ACTOR	App Web
DESCRIPCIÓN	La App Web solicita al modelo registrar el perfil de usuario.

CASO DE USO	Actualizar Perfil de Usuario
ACTOR	App Web
DESCRIPCIÓN	La App Web solicita al modelo actualizar el perfil de usuario.

CASO DE USO	Guardar Perfil de Usuario
ACTOR	App Web
DESCRIPCIÓN	La App Web solicita al modelo salvar el estado del perfil de usuario.
CASO DE USO	Cargar Perfil de Usuario
ACTOR	App Web
DESCRIPCIÓN	La App Web solicita al modelo recuperar el perfil de usuario almacenado en la base de datos.
CASO DE USO	Expandir Consulta
ACTOR	App Web
DESCRIPCIÓN	La App Web solicita al modelo realizar la expansión de una consulta.
CASO DE USO	Reclasificar Documentos
ACTOR	App Web
DESCRIPCIÓN	La App Web solicita al modelo reclasificar los documentos recuperados dada determinada consulta.

Tabla 13. Casos de Uso en Formato Compacto

4.2 FASE DE ELABORACIÓN

En esta fase se busca definir la arquitectura del sistema y realizar el modelado de la misma.

4.2.1 Arquitectura del Modelo

Para la implementación de MOBIRSE, se optó por definir una arquitectura multicapa [83], puesto que esta permite llevar a cabo el desarrollo en varios niveles, lo cual brinda ventajas en el proceso de construcción para la descomposición de actividades, flexibilidad y escalabilidad de la misma.

Teniendo en cuenta lo anterior, se definió una arquitectura, con una capa de acceso a datos y otra de lógica de negocio, tal como lo muestra la Figura 19.

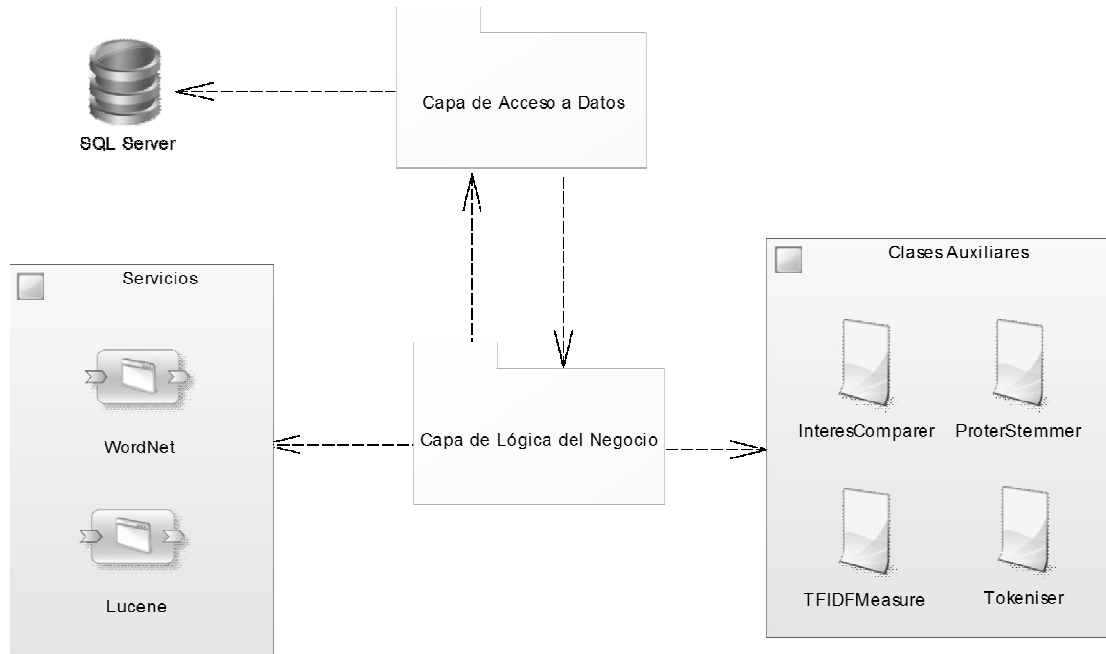


Figura 19. Arquitectura del Prototipo MOBIRSE

- **Capa Lógica de Negocio (CLN)**

En esta capa se reciben las peticiones provenientes de la aplicación Web de RSO y se le envían las respuestas tras el proceso. Se denomina capa de lógica del negocio porque aquí es donde se establecen todas las reglas que deben cumplirse. Así es como en esta capa se encuentran definidas las clases que soportarán la funcionalidad del modelo, más específicamente los módulos planteados en el modelo teórico, y clases auxiliares que serán utilizadas por las clases que conforman esta capa. Además estas clases hacen uso de los servicios Web WWordNet y WSLucene, los cuales son utilizados, para acceder al repositorio de términos de WordNet y para el procesamiento de texto respectivamente.

Cabe resaltar que esta capa es la que atiende directamente las solicitudes realizadas por la aplicación Web de RSO. Además también se comunica con la capa de datos para solicitar al gestor de la base de datos, almacenar o recuperar el estado del Perfil de Usuario Ontológico.

La capa de lógica del negocio es como se muestra en la Figura 20.

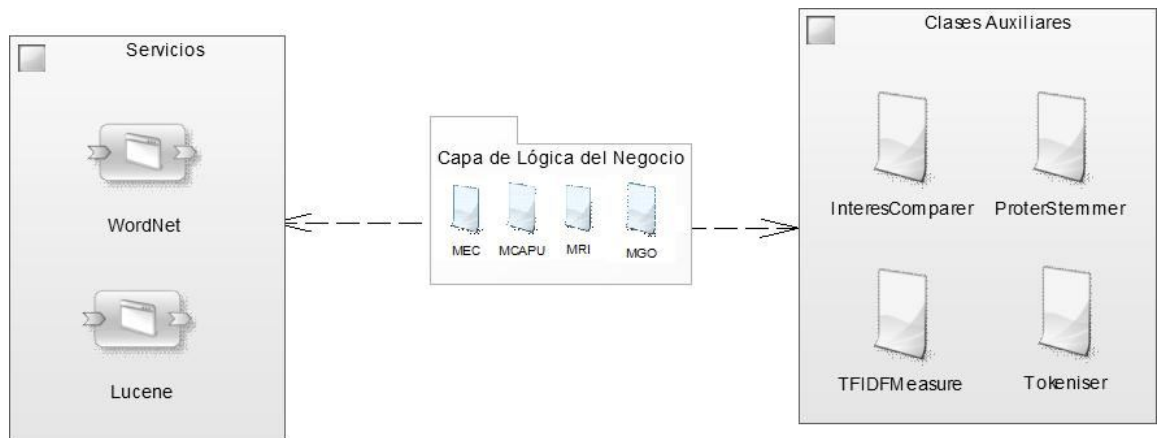


Figura 20. Capa Lógica del Negocio

- **Capa de Acceso a Datos**

En esta capa se encuentran definidas las clases que permitirán el acceso a datos, más específicamente, el acceso a la información referente al estado de la ontología que representa el perfil de usuario ontológico. Cabe aclarar que para salvar el estado de la ontología, se hace uso de un esquema de base de datos relacional, el cual es una réplica exacta de la ontología que representa el PU y funciona sobre el motor de base de datos SQL Server Express Edition 2008. La Figura 21 muestra la capa de acceso a datos.

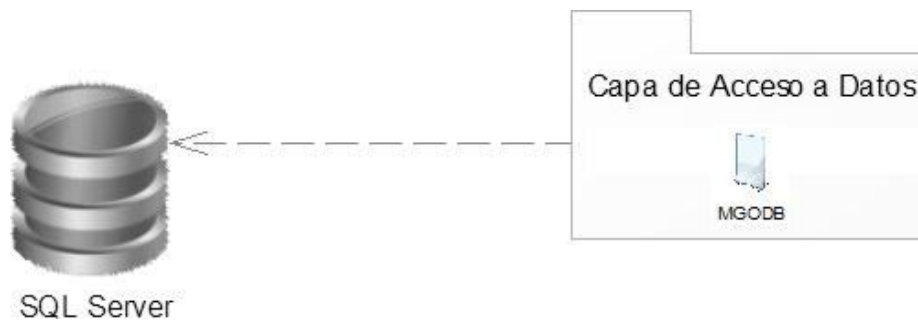


Figura 21. Capa de Acceso a Datos

La especificación de la arquitectura junto con los patrones de software usados se describe en detalle en el Anexo D.

4.3 FASE DE CONSTRUCCIÓN

En esta fase se definen en forma detallada los casos de uso, se realizan los diseños de los diagramas de clase, de secuencia, de despliegue y el modelo de base de datos los cuales, permiten observar el funcionamiento del prototipo a ser desarrollado.

4.3.1 Casos de Uso en Formato Extendido

Los casos de uso en formato extendido cobran mayor importancia en esta fase, puesto que son más detallados que los de formato de alto nivel. Este formato se utiliza para la especificación de requisitos más importantes o de mayor influencia en el funcionamiento del prototipo a desarrollar. Dado que estos casos de uso son mucho más extensos que los casos de uso de alto nivel, se describen con mayor detalle en el Anexo C.

4.3.2 Diagramas de Secuencia

Los diagramas de secuencia son utilizados para modelar la interacción entre objetos en una aplicación a través del tiempo y se modelan para cada caso de uso. La Figura 22 muestra uno de los diagramas de secuencia de las funcionalidades de MOBIRSE. Los diagramas de las demás funcionalidades se muestran en el Anexo C.

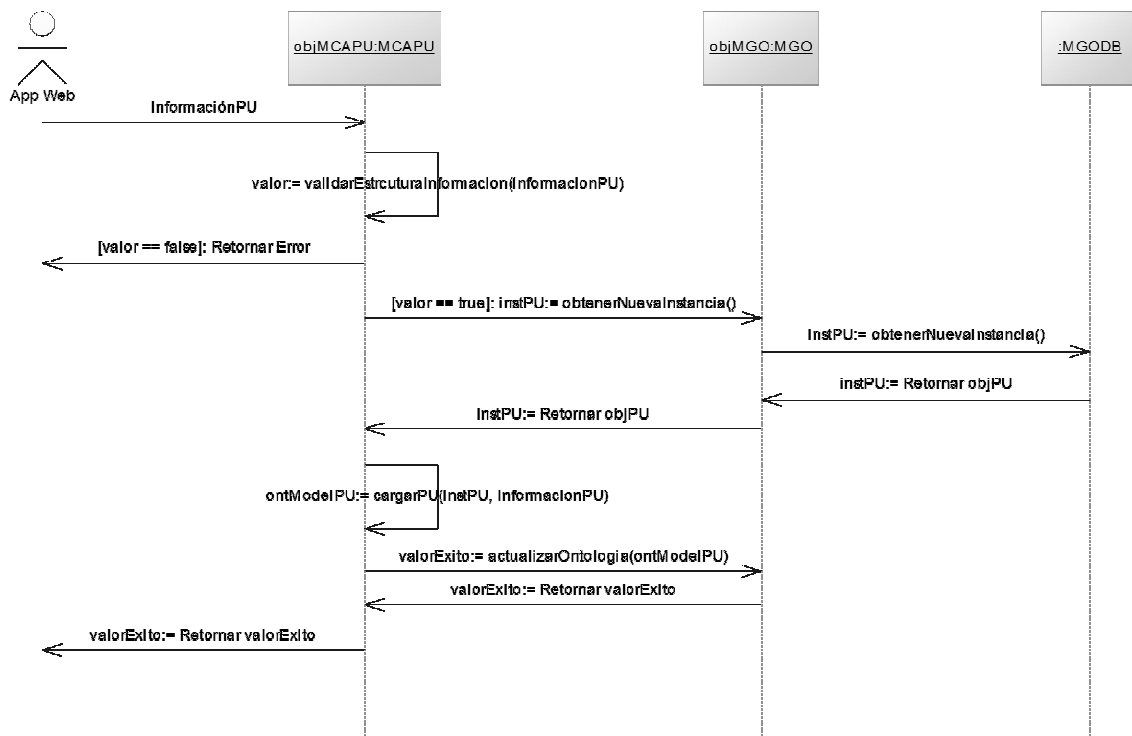


Figura 22. Diagrama de Secuencia Crear Perfil de Usuario

4.3.3 Diagramas de Clases

El diagrama de clases de la aplicación muestra las relaciones entre las clases necesarias para la construcción del prototipo. La Figura 23 muestra el diagrama general.

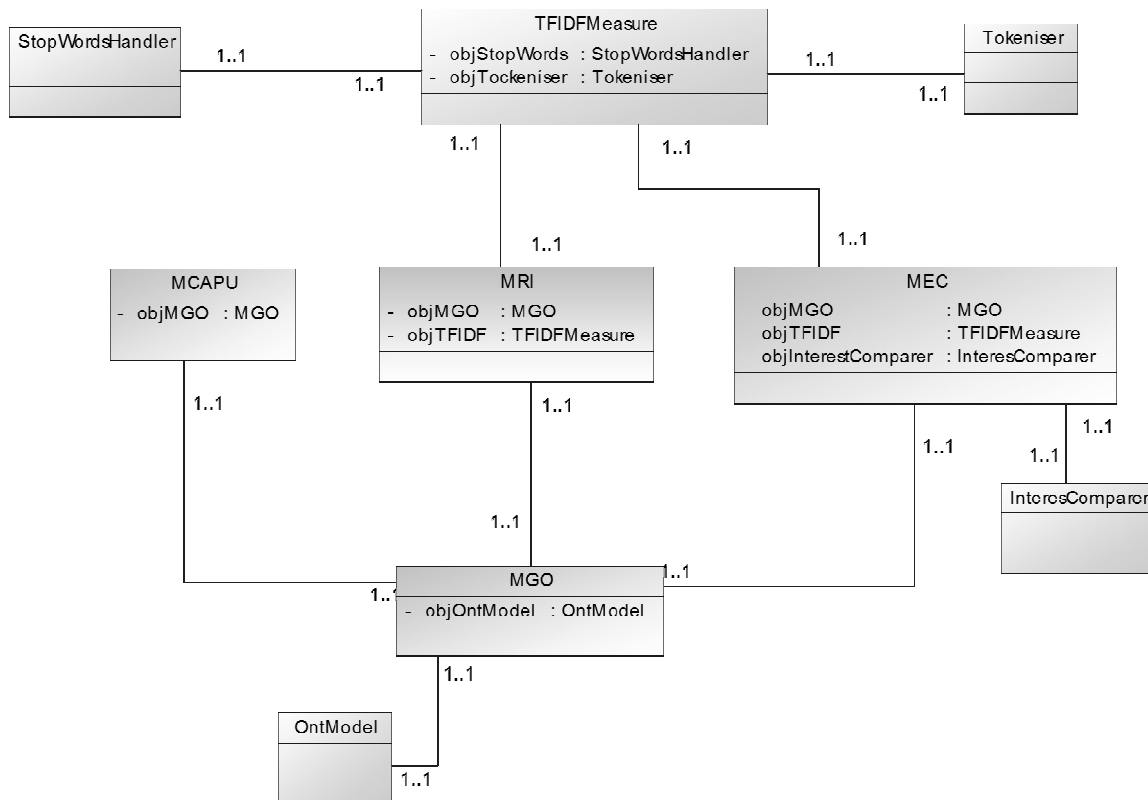


Figura 23. Diagrama de Clases MOIRSE

La Figura 23, muestra las clases MCAPU, MRI, MGO, MEC que representan los módulos del modelo planteado, las cuales ofrecen las funcionalidades principales descritas en el capítulo anterior. Además se tiene una serie de clases auxiliares como por ejemplo TFIDFMeasure, la cual es indispensable en el proceso de reclasificación, ya que se encarga de representar los documentos como vectores de términos ponderados. También, se observa la clase OntModel, la cual es muy importante puesto que a través de esta se logra la manipulación de la ontología.

4.3.4 Diagrama de Despliegue

El diagrama de despliegue contiene la forma como se muestran los componentes de la aplicación, teniendo en cuenta el servidor donde se implanta y los clientes que pueden acceder a su uso.

En la Figura 24, se puede apreciar que el primer nodo lo constituye el servidor Web que para el presente proyecto será IIS⁴⁴ (Internet Information Server), que es donde se encuentran las capas de acceso a datos y lógica del negocio del prototipo. En estas se encuentran las clases necesarias que dan soporte a los servicios que ofrece el modelo,

⁴⁴ ISS: es un servidor web y un conjunto de servicios para el sistema operativo Microsoft Windows.

como son: crear y actualizar el perfil de usuario ontológico, recuperar términos expandidos del perfil de usuario dada una consulta determinada y salvar el estado del perfil de usuario. También se alojan los servicios web WSWordNet y WSLucene, los cuales dan soporte para la identificación de términos y análisis léxico de texto respectivamente.

El segundo nodo es el correspondiente al servidor de base de datos, que aloja el motor de base de datos SQL Server, en el cual se almacena la información referente al perfil de usuario ontológico.

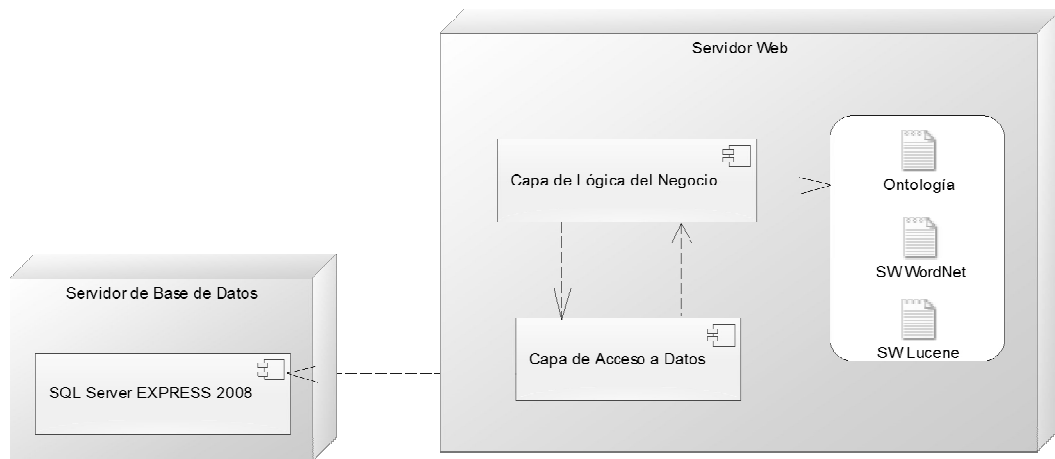


Figura 24. Diagrama de Despliegue de MOBIRSE

4.3.5 Modelo de Base de Datos

En la Figura 25, se muestra el modelo relacional de base de datos que representa el perfil de usuario, el cual es una réplica exacta de la ontología, la cual fue detallada en la Tabla 12.

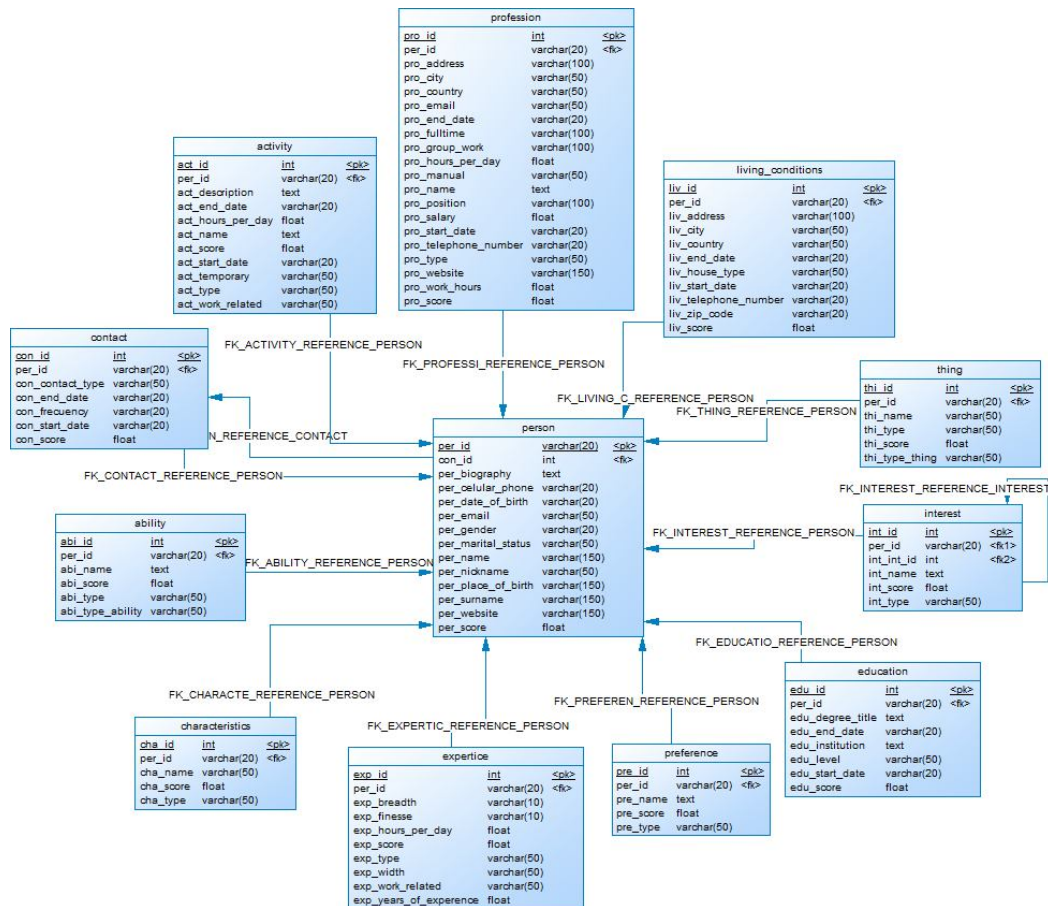


Figura 25. Modelo de Base de Datos de MOBIRSE

4.3.6 Desarrollo del Prototipo Software

4.3.6.1 Iteración 1

- Implementar el Módulo de Gestión de la Ontología (MGO), puesto que es el más complejo del modelo y se encarga de gestionar todo lo relacionado a la ontología que representa el perfil de usuario.
- Implementar el Módulo de Creación y Actualización del Perfil de Usuario (MCAPU), el cual ofrece funcionalidades concernientes a la creación, carga y actualización del perfil de usuario. Claro está, que para obtener la funcionalidad deseada de este módulo, se debe contar con los servicios implementados en el MGO.

Cabe mencionar que los dos módulos nombrados anteriormente, fueron desarrollados en paralelo, debido a que algunas funcionalidades del MCAPU hacen uso del MGO. Por otra parte, es importante resaltar que en esta iteración se realizaron las pruebas preliminares de funcionalidad, aplicando el diseño presentado en el Anexo D.

4.3.6.2 Iteración 2

- Implementar el Modulo de Expansión de Consulta (MEC), el cual tiene las funcionalidades correspondientes a la obtención de los conceptos similares a la consulta y que están basado en el perfil de usuario ontológico.
- Implementar el Modulo de Reclasificación de Información (MRI), el cual proveerá todas las funcionalidades entorno a la reclasificación de los resultados de búsqueda.

Cabe mencionar que tanto el MRI y el MEC son independientes, puesto que no dependen entre sí, sin embargo estos dos módulos, hacen uso del MGO, debido a que en ambos casos se tiene en cuenta el uso de la ontología. Por otra parte, es importante resaltar que en esta iteración se realizaron las pruebas preliminares de funcionalidad, aplicando el diseño presentado en el Anexo D.

4.3.6.3 Fase de Transición

En esta fase se debe tener en cuenta que el enfoque principal está ligado a las pruebas de validación del modelo, ahora bien, es importante aclarar que si bien al desarrollo del modelo, se le aplicaron las pruebas diseñadas en el Anexo D, su validación está estrechamente relacionada con la validación de la aplicación que haga uso de este modelo. Por esta razón la fase de transición del desarrollo de la aplicación web, la cual será descrita en detalle más adelante, complementara esta fase de transición.

4.3.6.4 API MOBIRSE

Como producto de la fase de implementación del modelo, se obtiene una API⁴⁵, que encapsula las funcionalidades del modelo. Dicha API debe ser referenciada por la aplicación Web que desee hacer uso del modelo. A continuación, la Tabla 14 muestra a modo de resumen las funcionalidades provistas por el API MOBIRSE.

Función	Modulo	Descripción	Entrada		Salida
			Nombre	Tipo	
loadSNUserProfile	MCAPU	Carga la información del Perfil de Usuario de la RSO a una instancia de la ontología.	infoUP	JsonObject	Estado de la operación (éxito, fallo).
			userId	string	
saveUserProfile	MGO	Almacena en esquema de base de datos el estado del perfil de usuario.	userId	string	Estado de la operación (éxito, fallo).
LoadOntologyModel	MCAPU	Carga la información del Perfil de Usuario desde la base de datos.	userId	string	Estado de la operación (éxito, fallo).

⁴⁵ API (Application Programming Interface): es un conjunto de funciones y procedimientos que ofrece cierta biblioteca para ser utilizado por otro software.

updateSNUserProfile	MCAPU	Se encarga de actualizar los pesos de la ontología.	terms	ArrayList	Estado de la operación (éxito, fallo).
			userId	string	
reRanking	MRI	Se encarga de reclasificar los resultados de búsqueda devueltos por la red social.	userId	string	ArrayList
			query	string	
			results	ArrayList	
getQueryExpansion	MEC	Retorna términos contenidos en el perfil de usuario y que guardan cierta similitud con la consulta enviada desde la App Web.	userId	string	ArrayList
			query	string	

Tabla 14. Funcionalidades API MOBIRSE

4.3.6.5 Plantilla de Aplicación del Modelo

Como resultado de la implementación del modelo, en la Tabla 15 se presenta una plantilla, la cual es una herramienta que puede ser utilizada por el desarrollador de la aplicación Web que funciona sobre la RSO específica, para utilizar el modelo de forma correcta y en consecuencia, obtener el máximo provecho de éste. Esta plantilla presenta una serie de pasos así como sus respectivas actividades. Es de precisar, que si no se cumple con alguna de las actividades del Paso 1, no se podrá aplicar el modelo sobre determinada RSO.

Pasos				
Paso	Actividad	Cumple		Observaciones y/o comentarios
Paso1	Elección de la RSO	Si	No	
	Identificar si posee API de desarrollo			
	Determinar si se tiene acceso a la información del PU a través de la API.			
	Determinar si se tiene acceso a la información de interacción (publicaciones, eventos, etiquetas, etc.) del usuario con la RSO a través de la API.			
	Determinar si se tiene acceso a los resultados de búsqueda a través de la API.			
Paso 2	Creación del PU			
	Recuperar los datos del PU y mapearlos al formato impuesto por el modelo.			
Paso 3	Actualización del PU			
	Definir la fuente de actualización del PU que será enviada al modelo.			
Paso 4	Expansión de Consulta			
	Definir el uso del servicio de expansión de consulta provisto por el modelo			
Paso 5	Reclasificación de Información.			
	Recuperar los resultados de búsqueda de la RSO, a través de la API, en base a la consulta dada.			
Paso 6	Evaluación de Relevancia de la Información Recuperada.			
	Consultas de prueba sobre la aplicación.			
	Evaluación de la aplicación utilizando Medidas de precisión			

Tabla 15. Plantilla para la Aplicación de MOBIRSE

5 IMPLEMENTACIÓN DE FACEBOOK META SEARCH ENGINE

Dado que la validación del modelo, se debe hacer a través de una aplicación Web que funcione sobre una RSO en particular, se optó por desarrollar una aplicación Web sobre la red social Facebook, que haga uso de la implementación del MOBIRSE, con los siguientes objetivos:

1. Validar el modelo propuesto.
2. Servir como ejemplo de uso del modelo propuesto.

Por lo anteriormente descrito, nace la idea de construir o desarrollar Facebook Meta Search Engine, el cual es un metabuscador que estará acoplado a la RSO Facebook

Vale la pena aclarar, que se seleccionó la red social Facebook como caso de estudio ya que fue la que se propuso en el anteproyecto, por las siguientes razones:

- Posee gran cantidad de información en su perfil de usuario y gran cantidad de recursos como publicaciones, grupos, eventos, entre otros, que serían una fuente de datos ideal para la creación y actualización del perfil de usuario.
- Posee un API de desarrollo bien definida, documentada y constantemente actualizada.
- Goza de mayor popularidad entre los usuarios de la Internet.

Además, se hace importante diligenciar la plantilla de aplicación del modelo producida en el capítulo anterior, a fin de determinar los elementos y acciones a seguir en el proceso de desarrollo de la aplicación y uso del modelo. La Tabla 16 muestra la plantilla de aplicación del modelo.

Pasos				
Paso	Actividad	Cumple		Observaciones y / o comentarios
Paso1	Elección de la RSO	Si	No	
	Identificar si posee API de desarrollo	X		
	Determinar si se tiene acceso a la información del PU a través de la API.	X		
	Determinar si se tiene acceso a la información de interacción (publicaciones, eventos, etiquetas, etc.) del usuario con la RSO a través de la API.	X		
	Determinar si se tiene acceso a los resultados de búsqueda a través de la API.	X		
Paso 2	Creación del PU			
	Recuperar los datos del PU y mapearlos al formato impuesto por el modelo.			Ya que para la creación del PU, el API MOBIRSE recibe un array asociativo del tipo

		JsonObject, es necesario que la información del PU de Facebook, sea mapeada en un array de este tipo.
Paso 3	Actualización del PU	
	Definir la fuente de actualización del PU que será enviada al modelo.	Dado que el modelo restringe que los recursos que pueden ser utilizados para la actualización del PU, es texto plano, se definió que las fuentes de actualización serán las publicaciones del muro y las consultas realizadas.
Paso 4	Expansión de Consulta	
	Definir el uso del servicio de expansión de consulta provisto por el modelo	Tras cada consulta realizada se debe invocar este servicio al API MOBIRSE, el cual de existir conceptos relacionados con la consulta ingresada retomara una colección de términos. Ahora bien, se deberá decidir si se hace un procesamiento extra de estos términos a fin de redefinir la consulta digitada. Para este caso se decidió mostrar los términos devueltos por el modelo, a modo de autocompletado.
Paso 5	Reclasificación de Información.	
	Recuperar los resultados de búsqueda de la RSO, a través de la API, en base a la consulta dada.	Una vez que se recuperan los resultados de búsqueda mediante la API de Facebook, estos serán enviados al módulo de reclasificación del API MOBIRSE, para que haga la respectiva reclasificación.
Paso 6	Evaluación de Relevancia de la Información Recuperada.	
	Consultas de prueba sobre la aplicación.	Para probar la aplicación se realizaran pruebas con usuarios reales afiliados a la Red Social Facebook.
	Evaluación de la aplicación utilizando Medidas de precisión	Para medir la precisión del modelo propuesto y en consecuencia de la aplicación desarrollada, se plantea calcular las siguientes medidas de precisión: Estadísticas Kappa y precisión at K.

Tabla 16. Plantilla de Aplicación de MOBIRSE sobre FMSE

Teniendo en cuenta lo anteriormente descrito, a continuación se detallan cada una de las fases necesarias para el proceso de creación de la aplicación.

5.1 FASE DE INICIO

En esta fase se definen los casos de uso más relevantes a un alto nivel con su respectivo diagrama de casos de uso, con el objetivo de establecer los requisitos mínimos para el desarrollo del prototipo. Posteriormente se identifica y define, a modo general una arquitectura potencial que permita guiar el proceso de desarrollo.

En general el proceso de desarrollo se llevo a cabo utilizando la herramienta de desarrollo Visual Studio 2010 con Framework 3.5 de Microsoft .NET, utilizando el lenguaje de programación C# y motor de base de datos SQL Server Express Edition 2008.

Igualmente también se hizo necesario contar con los siguientes elementos:

- **APIS para interactuar con la RSO Facebook:** se opta por utilizar Facebook C# SDK [84], la cual es una API de desarrollo para la plataforma Facebook, que permite extraer la información contenida en dicha red social.
- **Librerías para el trabajo con Ajax sobre la plataforma .NET:** se hará uso del ASP.NET AJAX Control Toolkit [85], para realizar peticiones asíncronas al servidor web.
- **Librerías para el trabajo con Java Script:** se selecciono el framework Jquery [86] para todo lo relacionado con Javascript⁴⁶ y Ajax⁴⁷[87], puesto que es uno de los que goza de mayor popularidad, basto numero de plugins⁴⁸ y suficiente documentación en internet.

5.1.1 Análisis de Requerimientos

Se desarrollará una aplicación Web sobre la red social Facebook, que permita realizar búsquedas de acuerdo a las necesidades de información de los usuarios afiliados a dicha red social, y que cumpla con los siguientes requisitos:

- Permitir al usuario el registro en la aplicación.
- Permitir al usuario el acceso a la aplicación.
- Permitir realizar consultas y los resultados de búsqueda devueltos, deberán presentar una precisión aceptable en las páginas que serán desplegadas al usuario.
- Mostrar sugerencias de búsqueda si la consulta dada está relacionada con los intereses que conforman el perfil de usuario.
- Permitir salir al usuario de la aplicación y en consecuencia salvar el estado del perfil de usuario.
- La interfaz grafica que se presente al usuario deberá ser amigable y de fácil interacción con el mismo.

⁴⁶ JavaScript: lenguaje de programación del lado del cliente, puesto que es el navegador web el encargado de interpretar sus instrucciones. Permite la manipulación de los elementos que componen una página HTML, como botones, tablas, etc.

⁴⁷ Ajax: acrónimo de Asynchronous Java Script And XML (Java Script Asíncrono y XML), es una técnica de desarrollo web, que permite a las páginas web intercambiar datos con el servidor, sin necesidad de recargar toda la pagina tras cada petición.

⁴⁸ Plugin: un plugin es un trozo de código software que se instala o se agrega a un sistema más grande, para añadirle a éste último alguna funcionalidad específica.

5.1.2 Diagramas de Casos de Uso

Teniendo en cuenta lo mencionado en la sección 4.1.2, la Figura 26, muestra el diagrama de casos de uso de Facebook Meta Search Engine.

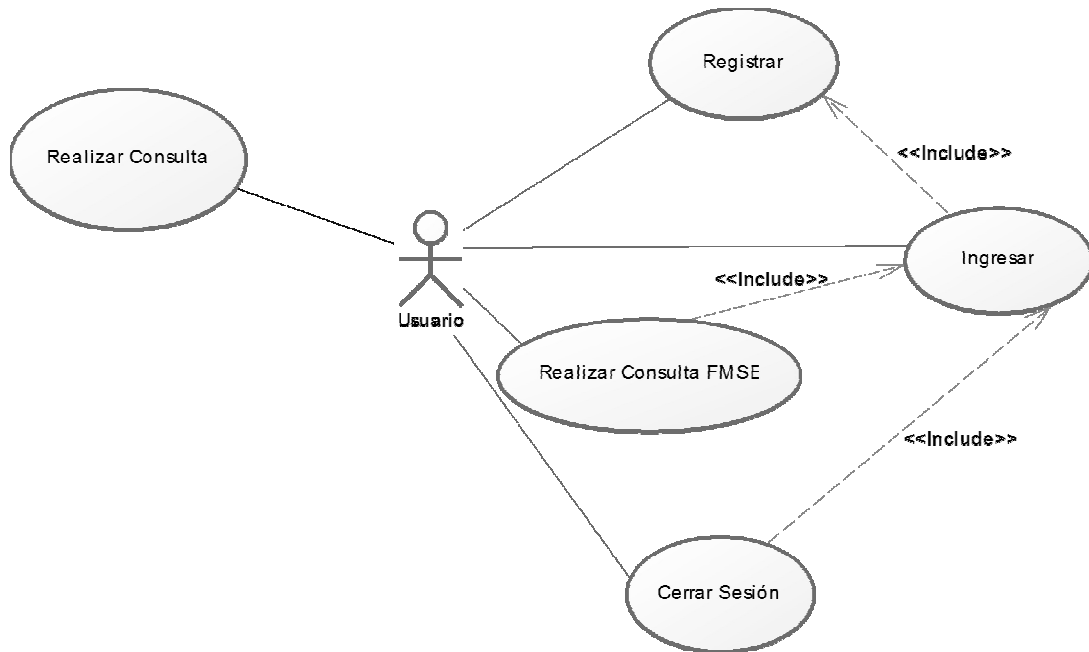


Figura 26. Diagrama Casos de Uso Facebook Meta Search Engine

5.1.3 Casos de Uso en Formato Compacto

Teniendo en cuenta lo mencionado en la sección 4.1.3, la Tabla 17 muestra los casos de uso en formato compacto.

CASO DE USO	Registrar
ACTOR	Usuario
DESCRIPCIÓN	El usuario desea registrarse en la aplicación para hacer uso de esta.

CASO DE USO	Ingresar
ACTOR	Usuario
DESCRIPCIÓN	El usuario desea ingresar o iniciar sesión para hacer uso de la aplicación.

CASO DE USO	Realizar Consulta
ACTOR	Usuario
DESCRIPCIÓN	El usuario desea realizar una consulta sobre el buscador web de la aplicación.

CASO DE USO	Cerrar Sesión
ACTOR	Usuario
DESCRIPCIÓN	El usuario desea cerrar sesión para salir de la aplicación.

Tabla 17. Casos de Uso en Formato Compacto

5.2 FASE DE ELABORACIÓN

En esta fase se busca definir la arquitectura del sistema y realizar el modelado de la misma.

5.2.1 Arquitectura de la Aplicación

En el siguiente apartado, se muestra el diseño arquitectónico de la aplicación web, la cual está basada en una arquitectura multicapa [83]. Para los propósitos de la aplicación se optó por definir una arquitectura tres capas, tal como lo muestra la Figura 27.

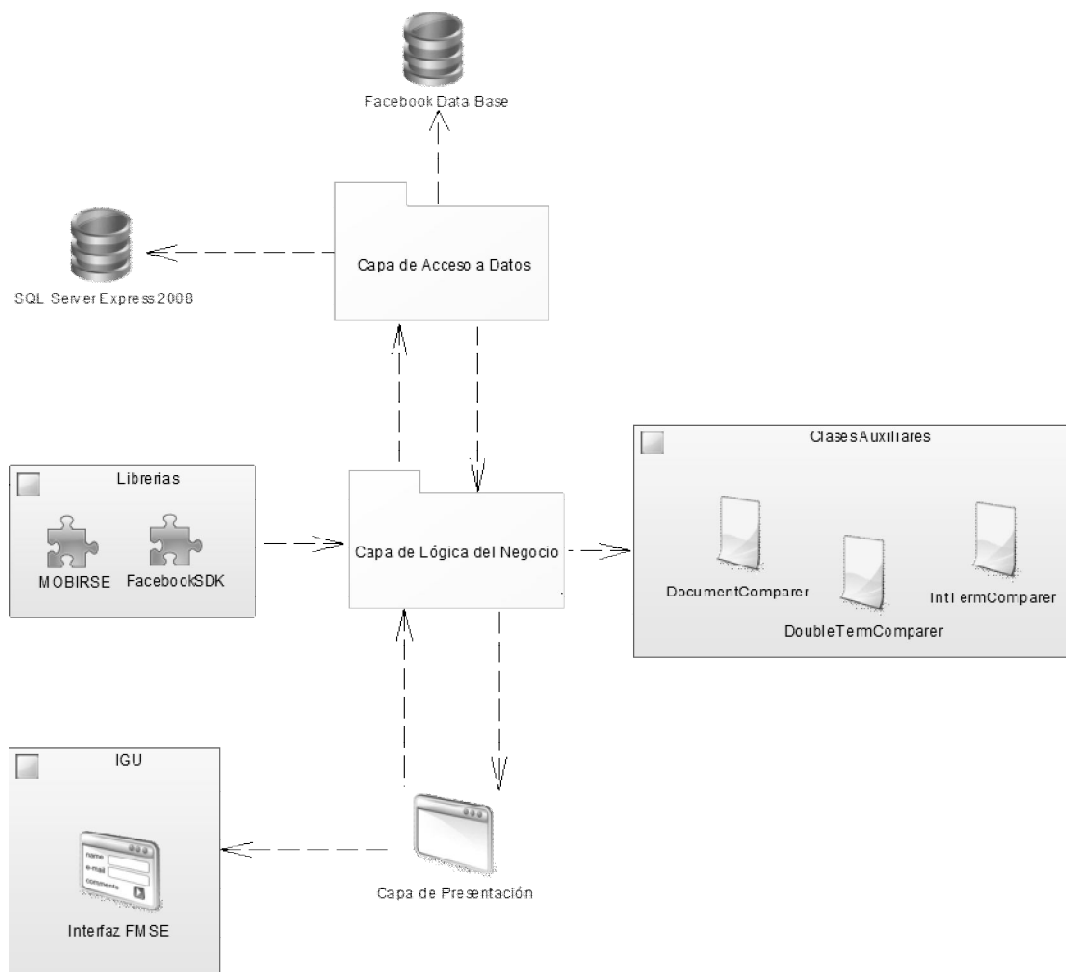


Figura 27. Arquitectura tres capas del prototipo Facebook Meta Search Engine

- **Capa de presentación**

En esta capa se encuentran los formularios web que recogen las solicitudes del usuario y las envían a la capa de lógica de negocio.

De este modo, fue necesario diseñar la interfaz del buscador, que es el medio por el cual el usuario interactúa con la aplicación. Para lograr un diseño adecuado de la interfaz, se tuvieron en cuenta los siguientes aspectos:

- Las fuentes y colores de los elementos de la interfaz deberían ser similares a los mostrados por Facebook.
- Ventanas de confirmación, ayuda e información y mensajes de espera deberían ser similares a los mostrados por Facebook.

- Las operaciones de registro, acceso y cierre de sesión, deberían hacerse mediante llamadas asíncronas para evitar que el usuario vaya de una página otra.

Además, debería diseñarse un logo que fuera representativo, y que involucrara los elementos que hacen parte de la propuesta, tanto del modelo como de la aplicación web. Tales elementos son Perfil de usuario, ontologías, recuperación de información y Redes Sociales, este ultimo como eje central, puesto que gracias a la información que provee Facebook, el modelo planteado cobra vida.

Teniendo en cuenta lo anterior, se diseño una sola interfaz que se encargue de capturar las solicitudes, registro, acceso, consulta y cierre de sesión. La Figura 28 muestra la interfaz de la aplicación acoplada a la red social Facebook.



Figura 28. Interfaz de FMSE

La Figura 29 muestra la interfaz de la aplicación tras una operación de búsqueda.



Figura 29. Interfaz de FMSE tras una operación de búsqueda

- **Capa de Lógica de Negocio**

Esta capa contiene las clases encargadas de procesar las operaciones realizadas por el usuario a través de la capa de presentación. Es en esta capa, donde la aplicación Web hace uso de los servicios contenidos por el API MOBIRSE para dar soporte a las operaciones de: registro del perfil de usuario, acceso a la aplicación, realizar y expandir consulta, salir de la aplicación. Igualmente las clases aquí contenidas tienen también la lógica necesaria, para determinar los llamados al modelo para realizar la actualización del perfil de usuario, más específicamente, dichos llamados se realizan cuando el usuario: se registra, accede, sale y realiza una consulta.

También esta capa se comunica con la capa de acceso de a datos para consultar o actualizar la última publicación realizada en el muro del usuario, e igualmente para extraer información concerniente al perfil de usuario, publicaciones y resultados de búsqueda de Facebook.

La Figura 30, muestra la capa de lógica de negocio de la aplicación.



Figura 30. Capa de Lógica de Negocio

- **Capa de Acceso a Datos**

Esta capa se comunica con la de lógica de negocio, y en esta se encuentran las clases, que dan soporte a operaciones que se quieran realizar sobre la base datos de SQL Server o sobre la de Facebook. La comunicación con esta última, se logra mediante el API de desarrollo Facebook C# SDK, la cual permite recuperar información de la Red Social Facebook, referente a:

- Información del Perfil de Usuario.
- Publicaciones, eventos, grupos, etc.
- Resultados de Búsqueda.

La Figura 31, muestra la capa de Acceso a Datos.

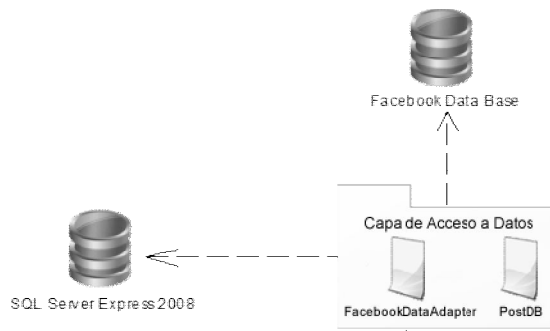


Figura 31. Capa de Acceso a Datos

5.3 FASE DE CONSTRUCCIÓN

5.3.1 Casos de Uso en Formato Extendido

Los casos de uso en formato extendido cobran mayor importancia en esta fase, puesto que son más detallados que los de formato de alto nivel. Este formato se utiliza para la especificación de requisitos más importantes o de mayor influencia en el funcionamiento del prototipo a desarrollar. Dado que estos casos de uso son mucho más extensos que los casos de uso de alto nivel, se describen con mayor detalle en el Anexo E.

5.3.2 Diagramas de Secuencia

Los diagramas de secuencia son utilizados para modelar la interacción entre objetos en una aplicación a través del tiempo y se modelan para cada caso de uso. La Figura 32 muestra uno de los diagramas de secuencia de las funcionalidades de Facebook Meta Search Engine. Los diagramas de las demás funcionalidades se muestran en el anexo E.

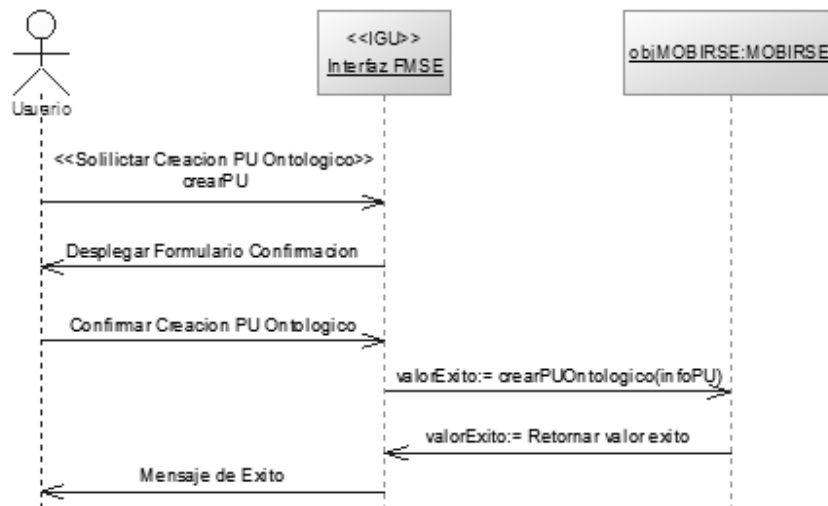


Figura 32. Registrar Perfil de Usuario

5.3.3 Diagrama de Clases

El diagrama de clases de la aplicación muestra las relaciones entre las clases necesarias para la construcción del prototipo FMSE. La Figura 33, muestra el diagrama de clases de la aplicación, en el cual se observan las clases que dan soporte a la funcionalidad del metabuscador. La clase Search funciona como un controlador y captura los eventos provenientes de la interfaz Web. La clase FacebookPDO, es la encargada de interactuar directamente con el API MOBIRSE y transformar los datos del Perfil de Usuario de Facebook al formato requerido por dicha API y otras funcionalidades propias de la aplicación. Además se cuenta con algunas clases auxiliares como DocumentComparer, DoubleTermComparer, IntComparer y PostManager las cuales son utilizadas para complementar el trabajo de las clases principales.

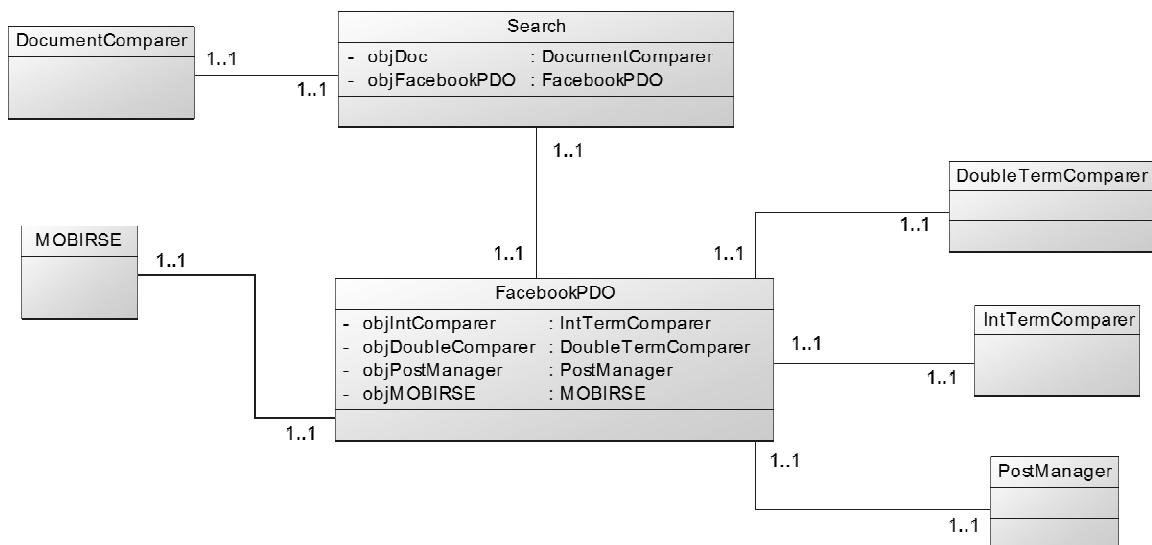


Figura 33. Diagrama de Clases de Facebook Meta Search Engine

5.3.4 Diagrama de Despliegue

El diagrama de despliegue contiene la forma como se muestran los componentes de la aplicación, teniendo en cuenta el servidor donde se implanta y los clientes que pueden acceder a su uso. La Figura 34 muestra el diagrama de despliegue de la aplicación.

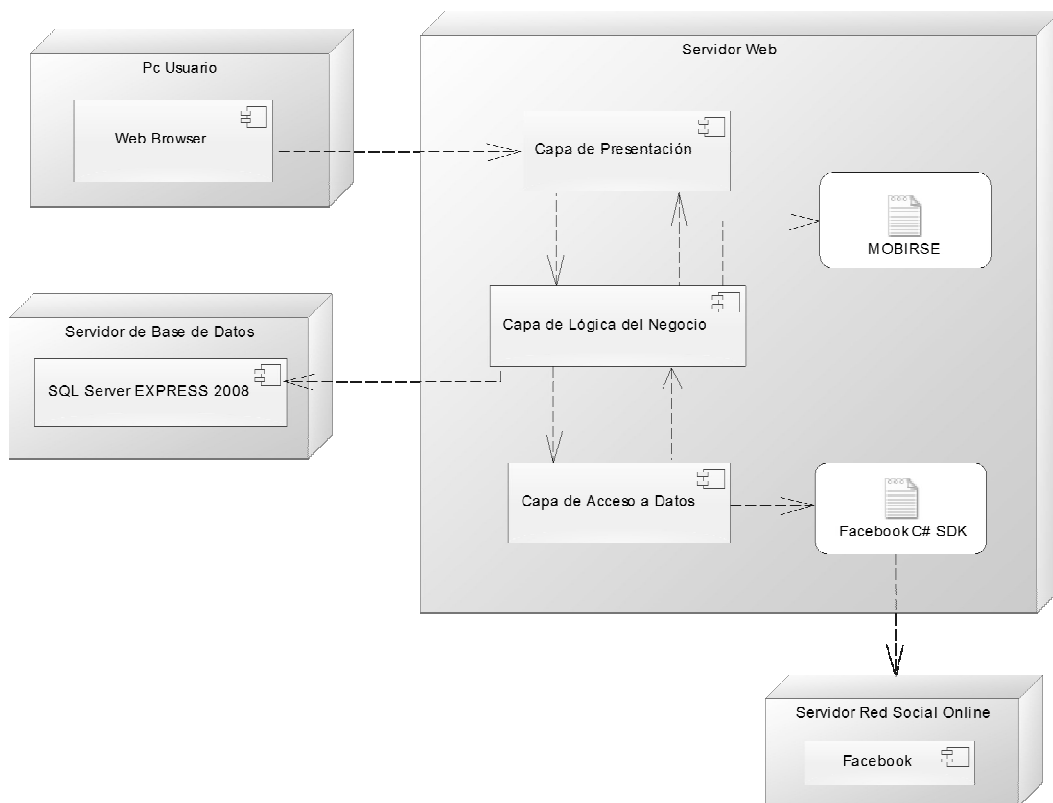


Figura 34. Diagrama de Despliegue FMSE

El primer nodo representa el equipo del usuario que accede a la aplicación a través de un navegador Web.

El segundo nodo representa el servidor web (Internet Information Services), el cual aloja las capas de la arquitectura mostradas en la Figura 27.

El tercer nodo lo representa el servidor de base de datos, el cual aloja el motor de base de datos de SQL Server, en el cual se almacena la información del perfil de usuario que necesita ser procesada por el API MOBIRSE y demás información propia de la lógica de la aplicación web.

El cuarto nodo hace referencia al servidor donde se aloja la información de la Red Social Facebook.

5.3.5 Modelo de Base de Datos

En la Figura 35, se muestra el modelo relacional de la Base de Datos, el cual consta de una única tabla, la cual es utilizada para determinar si hay nuevas publicaciones disponibles, las cuales son utilizadas para poder llevar a cabo el proceso de actualización del perfil de usuario de cada usuario registrado en la aplicación. Cabe aclarar que si bien se hubiera podido utilizar cualquier otro tipo de almacenamiento (archivos xml, texto plano, etc.), en lugar de un esquema

relacional para lograr este cometido, se opto por utilizar el motor de base de datos de sql server, ya que si el número de usuarios registrados en la aplicación crecía, el motor de base de datos mantendría el rendimiento de la aplicación. La Tabla 18, muestra los detalles de la tabla en cuestión.

PostDB	
PER_ID	varchar(20) <pk>
DATA_LAST_POST	datetime
POST_ID	varchar(50)

Figura 35. Modelo de Base de Datos FMSE

Campo	Descripción
PER_ID	Representa el id de cada usuario registrado en la aplicación.
DATA_LAST_POST	Representa la fecha de la última publicación.
POST_ID	Representa el id de la última publicación realizada.

Tabla 18. Descripción de la Tabla PostDB

5.3.6 Desarrollo del Prototipo Software

5.3.6.1 Iteración 1

- Implementar la capa de acceso a datos de la RSO, es decir, desarrollar las funcionalidades que permiten acceder a la información del perfil de usuario, a las publicaciones del muro y a los resultados de búsqueda de Facebook.
- Implementar las clases necesarias en la capa de lógica de negocio que permitan realizar la adaptación de la información del perfil de Usuario de Facebook, al formato requerido por el API MOBIRSE.
- Desarrolla la interfaz del prototipo software, para ello se tiene en cuenta la familiarización de un usuario típico con la interfaz de un buscador Web. Por esta razón las interfaces de usuario están orientadas hacia el estándar manejado en una interfaz gráfica tradicional de búsqueda, es decir una caja de texto en la que el usuario introduce su consulta, un botón que inicia el proceso de búsqueda y un logo distintivo con el nombre del buscador, esto fue mostrado en la Figura 28, al describir la capa de presentación de la arquitectura tres capas. En esta misma interfaz se muestran además los resultados de búsqueda con el título, el tipo y una descripción corta de cada resultado.

5.3.6.2 Iteración 2

- Implementar la funcionalidad del prototipo software, el cual hace uso del API MOBIRSE desarrollada previamente y se acopla el metabuscador implementado a la red social Facebook. Además se realizan las pruebas preliminares del mismo. Para esto se tienen en cuenta las herramientas mencionadas al inicio de este

capítulo y los artefactos resultantes de las anteriores fases. Las pruebas se describen con mayor detalle en el Anexo F.

5.4 FASE DE TRANSICIÓN

Esta fase tiene como enfoque principal, las pruebas de validación del prototipo, las cuales se encargan de validar el prototipo software desarrollado, según las expectativas del usuario. Para tal fin se entrenaron o capacitaron a los usuarios en el uso del sistema con el objetivo de ajustar factores como la usabilidad y el funcionamiento general del sistema. Con base en los resultados proporcionados por dichas pruebas, se realizan las correspondientes correcciones y modificaciones al sistema.

Las pruebas de rendimiento fueron aplicadas a 5 usuarios afiliados a la Red Social Facebook, de tal manera que cada uno de ellos realizó las tareas de: registro, acceso, búsqueda y cierre de sesión, para los cuales sus valores promedios de tiempo (segundos) fueron los mostrados en la Tabla 19, para mayor información ver Anexo F.

Tarea	Tiempo Promedio de Ejecución (segundos)
Registro del perfil de usuario	7.2
Acceso a la aplicación	4.2
Búsqueda	2.2
Cerrar sesión	4.4

Tabla 19. Tiempos de Respuesta de FMSE

Las pruebas de usabilidad fueron realizadas a estudiantes de últimos semestres del programa de Ingeniería de Sistemas de la Universidad del Cauca, los cuales accedieron voluntariamente a hacer las pruebas de la aplicación, de tal manera que cada estudiante realizó una prueba de usabilidad (Ver Anexo F para más información), donde se refleja su grado de satisfacción con la aplicación (flexibilidad, diseño, ayuda). Se tuvieron en cuenta los aspectos:

- Visibilidad del estado del sistema
- Relación entre sistema y mundo real
- Consistencia y estándares
- Reconocer en lugar de recordar (reconocimiento del sitio donde se encuentran)
- Ayuda y documentación
- ¿Cómo califica globalmente el sitio Web analizado?

Para cada aspecto se realizó una serie de preguntas a las que se debía responder, según el grado de satisfacción del evaluador, como: Excelente, Bueno, Neutro, Regular, Deficiente. Las preguntas y respuestas se pueden ver en detalle en el Anexo F. Al

terminar la prueba se realiza la ponderación de resultados que se muestra en la Figura 36.

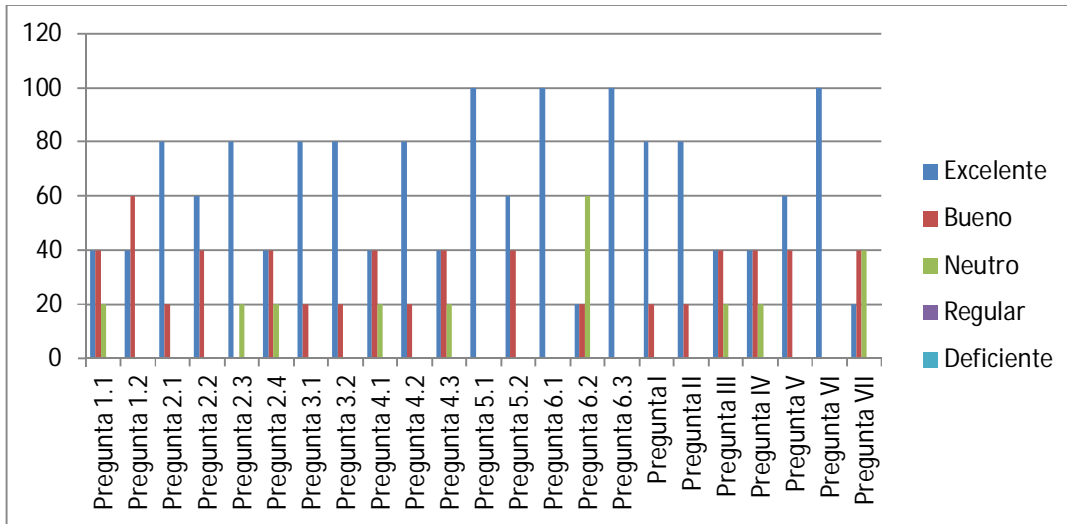


Figura 36. Resultados del test de usabilidad

Como se observa, los mayores resultados de calificación de la aplicación corresponden a los porcentajes de excelente y bueno, con lo que se concluye que los resultados son satisfactorios para los usuarios en cuanto a la usabilidad y al funcionamiento general del sistema. Los porcentajes y detalles de los resultados, así como las preguntas realizadas, se pueden consultar en el Anexo F.

6 VALIDACIÓN DEL PROTOTIPO

Para los sistemas de recuperación de información, cobra gran importancia medir la efectividad del proceso de recuperación, lo cual se logra evaluando la relevancia de los documentos recuperados. Ahora bien, medir la relevancia es algo subjetivo, ya que lo que para una persona puede ser relevante, para otra puede que no. Para dar cumplimiento al proceso de validación, se definieron dos fases: Pruebas Alfa y Pruebas Beta, las cuales serán detalladas a continuación:

6.1 Pruebas Alfa

Es necesario precisar, que son aquellas que se realizan en un entorno controlado y son guiadas por los desarrolladores de la aplicación que se pretende evaluar [88]. En este punto es importante medir la relevancia de la información recuperada por el meta buscador (FMSE) implementado, para abordar esta cuestión, se han definido ciertas métricas encaminadas a determinar la relevancia de un sistema de recuperación de información, entre ellas están la precisión y la precisión en los primeros k resultados [55]. Además se hace necesario alguna medida que indique que tan de acuerdo están los evaluadores con la relevancia de los resultados recuperados por la aplicación. Para esto se hará uso del índice Kappa, más exactamente el índice Kappa de Fleiss [89]. Estas medidas serán utilizadas para evaluar el prototipo denominado Facebook Meta Search Engine (FMSE), presentado en el capítulo anterior.

A continuación se detallan: las características del proceso de evaluación, las medidas utilizadas y los resultados obtenidos.

6.1.1 CARACTERÍSTICAS DEL PROCESO DE EVALUACIÓN

La evaluación se realizó comparando los resultados de búsqueda arrojados por Facebook con los arrojados por FMSE. Para llevar a cabo lo anterior fue necesario considerar lo siguiente:

- En este proceso se contó con 10 usuarios pertenecientes al programa de Ingeniería de Sistemas, puesto que se debería contar con un grupo homogéneo para determinar el índice Kappa.
- Se les solicitó previamente a los usuarios, que modificaran su perfil de Facebook al idioma inglés, es decir, se les pidió que cambiaran sus gustos, preferencias y demás a este idioma. Igualmente se les solicitó que publicaciones realizadas en su muro (por ellos o mismos o por sus contactos) fueran también en el idioma inglés.
- Las consultas realizadas se muestran en la Tabla 20:

# Consulta	Texto de Consulta
1	Tools and technologies for software development
2	Software engineering
3	Consulta Abierta

Tabla 20. Consultas realizadas

- Se valuaran los primeros 10 resultados devueltos por cada consulta, ya que el estudio hecho por Thorsten et al [90], demuestra su validez para medir la relevancia en buscadores o meta buscadores.

Teniendo en cuenta lo anterior se procedió a calcular, las medidas que se mostraran a continuación:

6.1.2 PRECISIÓN

La precisión se puede definir como la proporción de material recuperado realmente relevante, del total de los documentos recuperados. Esta medida toma valores entre 0 y 1, donde valores cercanos a 1 representan un adecuado nivel en la precisión de la recuperación. Por consiguiente, la precisión está definida con siguiente ecuación:

$$\text{Precisión} = \frac{\text{Número de documentos relevantes recuperados}}{\text{Número total de documentos recuperados}}$$

6.1.2.1 Precisión en los Primeros K Resultados

Teniendo en cuenta el concepto de precisión definido anteriormente y las características de la evaluación, se procedió a calcular este índice.

Documento	Facebook		FMSE	
	% Exa	% Pre At -k	% Exa	% Pre At -k
1	60	60	100	100
2	70	65	100	100
3	70	66,66666667	100	100
4	50	62,5	90	97,5
5	60	62	80	94
6	50	60	70	90
7	40	57,14285714	40	82,8571429
8	30	53,75	20	75
9	10	48,88888889	10	67,7777778
10	30	47	10	62

Tabla 21. Resultados de Precisión en K = 10

$$\%Exactitud = \frac{\text{Nro de Personas que consideran el ítem como relevante}}{\text{Nro de Personas que realizan la prueba}} * 100$$

Ecuación 4

Vale la pena aclarar, que la columna (% Exa) indica la exactitud en cada uno de los documentos recuperados, la cual se define por (número de personas que consideraron el ítem como relevante, sobre el número de personas que realizaron la prueba), por 100, como se muestra en la Ecuación 4. La columna (% Pre At-k) muestra la precisión At-K en cada uno de los documentos recuperados en los dos sistemas evaluados.

De la Tabla 21, se pueden observar los resultados de la precisión en los K primeros resultados (K=10) para Facebook y FMSE. Estos resultados muestran:

1. Que Facebook tiene como punto máximo de exactitud un 70% y como punto mínimo un 10%, con lo cual se establece una precisión para K = 10 que oscila entre el 47% y el 66.6%.
2. Para FMSE se obtuvo un punto máximo de exactitud de 100% y un punto mínimo de 10%, con una precisión en K = 10 que oscila entre 62% y 100%.
3. Los 10 primeros resultados mostrados por FMSE, son más relevantes que los retornados por el buscador tradicional de Facebook.

La Figura 37, presenta la precisión en K documentos recuperados para los 2 sistemas comparados anteriormente.

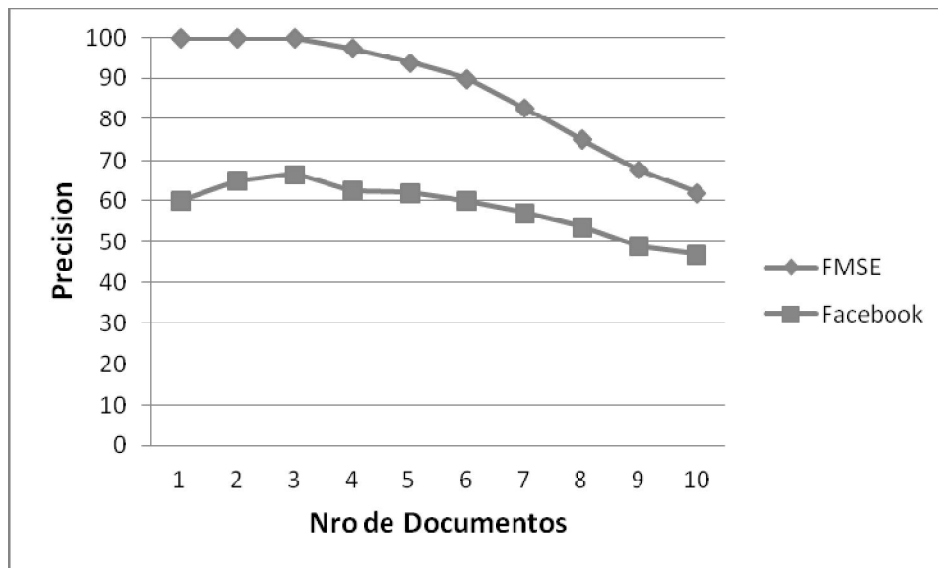


Figura 37. Precisión At K (K=10) para Facebook y FMSE

La Figura 37 muestra como resultado, la precisión promedio en diferentes valores de K, con resultados comprendidos entre el 62% y 100% para FMSE, mostrando así un alto

grado de precisión del modelo propuesto. Además en la gráfica se observa que los resultados de FMSE mantienen valores de precisión superiores a los que proporciona el buscador tradicional de Facebook para todos los valores de K.

6.1.3 ESTADÍSTICAS KAPPA

Para evaluar los resultados obtenidos por FMSE con el índice Kappa, se tuvo en cuenta que dos aspectos distintos entran a formar parte del estudio de fiabilidad: por una parte, la concordancia entre observadores, es decir, hasta qué punto los observadores no coinciden en su medición, y por otra parte el sesgo entre observadores definido como la tendencia de un observador a dar consistentemente valores mayores que el otro.

Dado lo anterior se utilizó el índice Kappa de Fleiss, el cual trabaja con cualquier número de observadores que proporcionan grados categóricos (en este caso R = Relevante, I = No relevante o Irrelevante), a un número fijo de documentos (en este caso de k = 1 hasta 10 documentos recuperados de la Web). Esta medida puede ser interpretada como el grado de concordancia observada entre varios observadores con una medida de consistencia la cual puede tomar valores entre -1 y +1. Mientras más cercano a +1, mayor es el grado de concordancia entre los observadores, por el contrario, mientras más cercano a -1, mayor es el grado de discordancia entre dichos observadores. El cálculo del índice Kappa se logra realizando el siguiente proceso:

1. Se calcula p_j , el cual representa la proporción de todos los jueces que se inclinaron por una determinada categoría, es decir, Relevante o No relevante.

$$p_j = \frac{1}{Nn} \sum_{i=1}^N n_{ij}$$

Ecuación 5

2. Se calcula P_i , medida que indica cuantas parejas juez-juez están de acuerdo, relativo al número de todos los posibles pares juez-juez.

$$P_i = \frac{1}{n(n-1)} \left[\left(\sum_{j=1}^k n_{ij}^2 \right) - n \right]$$

Ecuación 6

3. Se calcula \bar{P} , medida que representa la media de los P_i y la probabilidad de acuerdo entre dos jueces por azar.

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i$$

Ecuación 7

$$\bar{P}_e = \sum_{j=1}^k p_j^2$$

Ecuación 8

4. Finalmente, el cálculo del índice Kappa de Fleiss, esta dado por la siguiente ecuación:

$$Kappa_f = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

Ecuación 9

El cálculo del índice Kappa de Fleiss, fue calculado para los diez primeros documentos recuperados, calificados por 10 observadores, con base en el juicio de Relevancia y No Relevancia de cada documento recuperado.

En la Tabla 22 se muestra el juicio de los 10 observadores en los primeros 10 documentos recuperados para una misma consulta.

DOCUMENTO	OBSERVADORES (USUARIOS)									
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10
1	R	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R
4	R	R	R	NR	R	R	R	R	R	R
5	R	R	NR	R	R	R	R	R	R	NR
6	R	NR	R	R	R	R	NR	NR	R	R
7	NR	R	NR	R	R	NR	R	NR	NR	NR
8	R	NR	NR	NR	NR	NR	R	NR	NR	NR
9	NR	NR	NR	NR	NR	NR	NR	R	NR	NR
10	NR	NR	R	NR	NR	NR	NR	NR	NR	NR

Tabla 22. Apreciación de los observadores en k=10 resultados

En la Tabla 23, se puede observar que teniendo como base el primer documento (K=1), el 100% de los observadores (jueces) lo consideran relevante. Cuando se evalúan los 10 primeros documentos (K=10) se tiene que un 62% de los jueces considera los documentos como **Relevantes** mientras que el restante 38% los considera como **No Relevantes**. El índice Kappa para los primeros 10 documentos muestra que el número de personas que están de acuerdo es considerable y se ubica dentro de una concordancia moderada de acuerdo a los rangos establecidos por Landis y Koch [91].

Documento	Total: R	Total: NR	Pi
1	10	0	1
2	10	0	1
3	10	0	1
4	9	1	0,8
5	8	2	0,64444444
6	7	3	0,53333333
7	4	6	0,46666667

8	2	8	0,64444444
9	1	9	0,8
10	1	10	0,8
Totales	62	38	7,68888889
p_i	0,62	0,38	
p_i^2	0,3844	0,1444	
\bar{P}	0,76888889		
\bar{P}_e	0,5288		
Kappa	0,5095265		

Tabla 23. Resultado del índice kappa en K=10

Cabe aclarar, que estas pruebas fueron realizadas para dos consultas más, obteniéndose buenos resultados con respecto a la concordancia de los observadores. Los detalles se encuentran consignados en el Anexo G.

6.2 Pruebas Beta

Son aquellas, que se llevan a cabo por los usuarios finales del producto software en los lugares de trabajo de los clientes. A diferencia de las pruebas alfa, el desarrollador no está presente. Así, la prueba beta es una aplicación en vivo del software en un entorno que no puede ser controlado por el desarrollador [88].

6.2.1 CARACTERÍSTICAS DEL PROCESO DE EVALUACIÓN

Para este proceso de evaluación, se pidió a 10 usuarios del programa de Ingeniería de Sistemas de la Universidad del Cauca, afiliados a la Red Social Facebook, que interactuaran con la aplicación, a su gusto y se les pidió que llenaran una encuesta que consta de las siguientes preguntas o interrogantes:

- **Pregunta 1:** ¿Qué tan fácil fue manejar la aplicación?
 - ✓ Difícil
 - ✓ Parcialmente Fácil
 - ✓ Fácil

- **Pregunta 2:** ¿El autocompletado ayudo con su búsqueda?
 - ✓ Si
 - ✓ No

- **Pregunta 3:** ¿Qué tan adecuado fue el tiempo de respuesta?
 - ✓ Malo
 - ✓ Aceptable
 - ✓ Bueno

- **Pregunta 4:** ¿Qué tan relevante fue la información obtenida?
 - ✓ Poco Relevante
 - ✓ Relevante
 - ✓ Muy Relevante

A continuación se muestra los resultados obtenidos, de este proceso de evaluación:

Qué tan fácil fue manejar la aplicación?

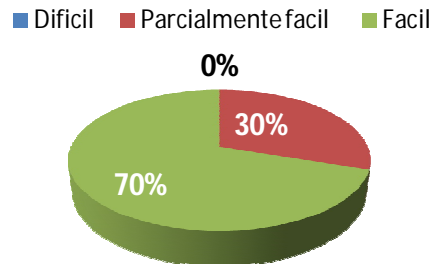


Figura 38. Resultados Pregunta Nro 1

Como se puede observar en la Figura 38, un 30% de los usuarios considera que el metabuscador es parcialmente fácil de usar, por el contrario ninguno de ellos considero que la aplicación fuera difícil de utilizar. Finalmente el 70% de los usuarios, consideran que la aplicación es fácil de usar, lo cual indica que el metabuscador es muy usable.

Qué tan adecuado fue el tiempo de respuesta?

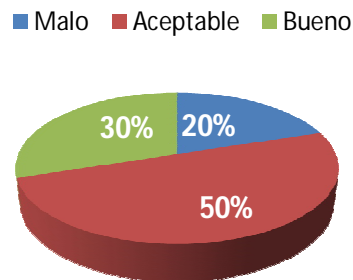


Figura 39. Resultados Pregunta Nro 2

Como se puede observar en la Figura 39, un 20% de los usuarios considero el tiempo de respuesta como malo, el 30% de ellos considero que era bueno y el 50% considero que el tiempo de respuesta es aceptable, lo cual nos lleva a concluir que el tiempo de respuesta promedio es aceptable.

El autocompletado ayudo con su búsqueda?

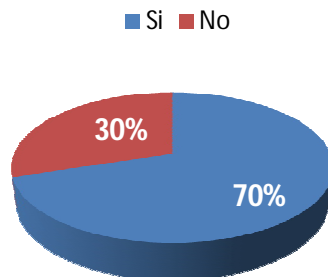


Figura 40. Resultados Pregunta Nro 3

Como se puede observar en la Figura 40, un 30% de los usuarios considero que el servicio de autocompletado no ayudo con su búsqueda, por el contrario, la gran mayoría, es decir el 70% de ellos considero útil este servicio, lo cual indica que el servicio de expansión de consulta provisto por el modelo es realmente útil.

Qué tan relevante fue la información obtenida?

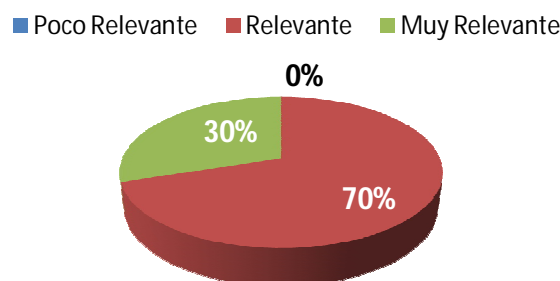


Figura 41. Resultados Pregunta Nro 4

Como se puede observar en la Figura 41, ninguno usuario considero que los resultados fueran poco relevantes, por otra parte, el 30% de los usuarios consideraron que los eran muy relevantes. Finalmente, la gran mayoría de los usuarios, es decir, el 70% de los usuarios consideraron como relevantes los resultados obtenidos, lo cual indica que el servicio de reclasificación provisto por el modelo es realmente útil.

Finalmente, se puede observar a nivel general, que los resultados obtenidos en las pruebas Beta respaldan los resultados obtenidos en las pruebas Alpha, y en consecuencia corroboran la utilidad y desempeño del modelo propuesto.

7 CONCLUSIONES, RECOMENDACIONES Y TRABAJO A FUTURO

Este capítulo describe inicialmente las principales conclusiones del trabajo realizado a las que se llegó durante su desarrollo, posteriormente presenta las recomendaciones, y finalmente propone los trabajos futuros.

7.1 CONCLUSIONES

- Se definió un modelo para la recuperación de Información en Redes Sociales Online, el cual se basa en la representación del Perfil de Usuario por medio de ontologías, a fin de mejorar la relevancia de la información recuperada en estos sitios Web.
- El modelo propuesto tienen como fortalezas que es adaptable a cualquier red social, además aporta una mejora a la precisión de los resultados de búsqueda, caso específico de Facebook, igualmente es fácil de usar y finalmente involucra dos temas importantes hoy en día como lo son las redes sociales y la web semántica. Sin embargo, presenta una serie de debilidades, puesto que funciona inicialmente solo para el idioma inglés y fue probado o aplicado únicamente a la red social Facebook.
- Se obtuvo un marco de referencia en lo concerniente a las Redes Sociales Online, puesto que se analizó la información del perfil de usuario, los servicios que ofrecen (principalmente el servicio de búsqueda), APIs de desarrollo y demás características presentes en estos sitios Web, generando así un referente teórico inexistente hasta este punto.
- La estrategia de implementar el modelo en forma de API, permite que este pueda ser agregado fácilmente como un módulo independiente por parte de la App Web que desee hacer uso del modelo, muestra de esto, es su uso en la implementación del metabuscador Facebook Meta Search Engine (FMSE).
- Los resultados experimentales muestran que el metabuscador implementado (FMSE) que hace uso del modelo propuesto, establece una mejora importante respecto a los mecanismos de recuperación de información utilizados por los buscadores tradicionales de las redes sociales, puesto que para el caso específico de Facebook, proporciona ventajas importantes tales como: un aumento en la recuperación de documentos relevantes y un mejor ranking de los mismos en el conjunto de resultados que son desplegados al usuario.

7.2 RECOMENDACIONES

- Si bien el modelo planteado, tiene definido una serie de servicios, que van encaminados a mejorar la relevancia de los documentos recuperados, el aprovechamiento del uso del modelo, depende en gran medida de que la aplicación Web que lo vaya a utilizar, lo haga teniendo en cuenta los pasos proporcionados en la plantilla de aplicación, puesto que la omisión en alguno de éstos, podría impactar de forma negativa en la relevancia de los documentos recuperados.

7.3 TRABAJO A FUTURO

- Incorporar la realimentación explícita, como mecanismo de actualización del perfil de usuario, puesto que el modelo planteado actualiza el perfil de usuario de forma implícita, es decir, sin intervención del usuario.
- Aplicar el modelo planteado a una red social de temática específica, como por ejemplo LinkedIn, a fin de observar el comportamiento del modelo sobre este tipo de redes sociales.
- Agregar un módulo, que permita obtener resultados de motores de búsqueda tradicionales como Google y Yahoo, con el objetivo de que el usuario disponga de una alternativa de búsqueda diferente a la proporcionada por la red social.

8 REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Ly, *Redes Sociales Virtuales: personas, sociedad y empresa*: ccee.edu.uy, 2009.
- [2] J. J. C. González, "Una aproximación a lo que son las necesidades de información," *Investig Bibliot*, 1991.
- [3] I. R. A. Sánchez, "La importancia social de la información," *Acimed*, 2001.
- [4] S. Arce, *et al.*, "Las comunidades virtuales y los portales como escenarios de gestión documental y difusión de información," *Anales de documentación*, 2001.
- [5] F. Javaloy, *et al.*, "Internet y movimientos sociales: un enfoque psicosocial," *Anuario de Psicología*, 2001.
- [6] Wikipedia. "Facebook", [en línea] <http://es.wikipedia.org/wiki/Facebook> [Consulta 04 de abril de 2011].
- [7] M. A. Abian, *El futuro de la Web: XML, RDF/RDFS, Ontologías y la Web Semántica*, 2005.
- [8] A. Sieg, *et al.*, "Web Search Personalization with Ontological User Profiles," *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 2007.
- [9] R. Baeza-Yates and B. Ribeiro-Neto, *Modern information retrieval*: simmons.edu, 1999.
- [10] A. N. Joinson, "Looking at, looking up or keeping up with people?: motives and use of facebook," *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 2008.
- [11] G. Barchini, *et al.*, "El rol de las ontologías en los SI," *Ingeniería informática*, 2007.
- [12] J. Contreras and J. Martínez-Comeche, *Ontologías: ontologías y recuperación de información*: recolecta.net, 2008.
- [13] O. D. L. ONTOLOGÍAS and S. L. WEB, "TUTORIAL ONTOLOGÍAS," *deki.uca.es*.
- [14] D. M. Sánchez, *et al.*, "The road toward ontologies," *Integrated Series in Information Systems*, 2006.
- [15] O. Corcho, *et al.*, "Methodologies, tools and languages for building ontologies. Where is their meeting point?," *Data & Knowledge Engineering*, 2003.
- [16] P. Muñoz, "Unos métodos en el desarrollo de ontologías," *titan.tel.uva.es*, 2000.
- [17] Y. Sure, *et al.*, "On-to-knowledge methodology (OTKM)," *Handbook on Ontologies, chapitre 6*, 2003.
- [18] M. Fernández and A. Gómez, "Overview and analysis of methodologies for building ontologies," *The Knowledge Engineering Review*, 2003.
- [19] M. Grüninger and M. S. Fox, "Methodology for the Design and Evaluation of Ontologies," *Proceedings of the Workshop*, 1995.
- [20] Protege. "Open Source Ontology Editor", [en línea] <http://protege.stanford.edu/> [Consulta 23 de junio de 2010].
- [21] J. A. C. P. Diego Fabian Gallego Fernandez, "Desarrollo de Ontologías para su Uso en Perfiles de Usuario en el entorno IMS," 2009.
- [22] S. Gauch, *et al.*, "User profiles for personalized information access," *Lecture Notes in Computer Science*, 2007.
- [23] G. Gentili, *et al.*, "Infoweb: An adaptive information filtering system for the cultural heritage domain," *Applied Artificial Intelligence*, 2003.
- [24] A. Moukas, "Amalthea information discovery and filtering using a multiagent evolving ecosystem," *Applied Artificial Intelligence*, 1997.

- [25] L. Chen and K. Sycara, "WebMate: a personal agent for browsing and searching," *AGENTS '98 Proceedings of the second international conference on Autonomous agents*, 1998.
- [26] D. Widyantoro, *et al.*, "Alipes: A swift messenger in cyberspace," *AAAI Technical Report SS-99-03. Compilation copyright © 1999, AAAI (www.aaai.org). All rights reserved*, 1999.
- [27] F. Asnicar and C. Tasso, "ifWeb: a prototype of user model-based intelligent agent for document filtering and navigation in the World Wide Web," *In Proc. of 6th International Conference on User Modelling (2-5 June 1997)* 1997.
- [28] A. Stefani and C. Strappavara, "Personalizing access to web sites: The SitelF project," *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia*, 1998.
- [29] A. Tan and C. Teo, "Learning user profiles for personalized information dissemination," *The 1998 IEEE International Joint Conference on 1998*.
- [30] A. Micarelli and F. Sciarrone, "Anatomy and empirical evaluation of an adaptive web-based information filtering system," *User Modeling and User-Adapted Interaction*, 2004.
- [31] A. Tan, "Adaptive resonance associative map," *Neural Networks*, 1995.
- [32] B. Sheth, *A learning approach to personalized information filtering*: Citeseer, 1994.
- [33] S. E. Middleton, *et al.*, "Capturing interest through inference and visualization: ontological user profiling in recommender systems," *Proceedings of the 2nd international conference on Knowledge capture* 2003.
- [34] A. Pretschner and S. Gauch, "Ontology based personalized search," *ictai*, 1999.
- [35] F. Liu, *et al.*, "Personalized web search by mapping user queries to categories," *Proceedings of the eleventh international conference on Information and knowledge management*, 2002.
- [36] C. Chen, *et al.*, "PVA: A self-adaptive personal view agent," *Journal of Intelligent Information Systems*, 2002.
- [37] P. Haase, *et al.*, "Collaborative and usage-driven evolution of personal ontologies," *The Semantic Web: Research and Applications*, 2005.
- [38] M. Golemati, *et al.*, "Creating an ontology for the user profile: Method and applications," *Proceedings of the First International Conference on Research Challenges in Information Science (RCIS)*, 2007.
- [39] M. Daoud, *et al.*, "A session based personalized search using an ontological user profile," *Proceedings of the 2009 ACM symposium on Applied Computing*, 2009.
- [40] S. He and M. Fang, "Ontological User Profiling on Personalized Recommendation in e-Commerce," *e-Business Engineering, 2008. ICEBE '08. IEEE International Conference on*, 2008.
- [41] A. Hernandez, "Guia sobre las Redes Sociales," 2007.
- [42] Facebook. "Red Social Facebook", [en linea] <http://www.facebook.com/> [Consulta 23 de junio de 2010].
- [43] MySpace. "Red Social MySpace", [en linea] <http://www.myspace.com/> [Consulta 23 de junio de 2010].
- [44] Twitter. "Red Social Twitter", [en linea] <http://twitter.com/> [Consulta 23 de junio de 2010].
- [45] Linkedin. "Red Social Linkedin", [en linea] <http://www.linkedin.com/> [Consulta 23 de junio de 2010].
- [46] Last.fm. "Red Social Last.fm", [en linea] <http://www.lastfm.es/> [Consulta 23 de junio de 2010].
- [47] Tagged. "Red Social Tagged", [en linea] <http://www.tagged.com/> [Consulta 23 de junio de 2010].
- [48] Hi5. "Red Social Hi5", [en linea] <http://www.hi5.com/> [Consulta 23 de junio de 2010].

- [49] Meetup. "Red Social Meetup", [en línea] <http://www.meetup.com/> [Consulta 23 de junio de 2010].
- [50] Bebo. "Red Social Bebo", [en línea] <http://www.bebo.com/> [Consulta 23 de junio de 2010].
- [51] MyLife. "Red Social MyLife", [en línea] <http://www.mylife.com/> [Consulta 23 de junio de 2010].
- [52] Friendster. "Red Social Friendster", [en línea] <http://www.friendster.com/> [Consulta 23 de junio de 2010].
- [53] MyHeritage. "Red Social MyHeritage", [en línea] <http://www.myheritage.es/> [Consulta 23 de junio de 2010].
- [54] Multiply. "Red Social Multiply", [en línea] <http://multiply.com/> [Consulta 23 de junio de 2010].
- [55] C. D. Manning, *et al.*, "An introduction to information retrieval," *dspace.cusat.ac.in*.
- [56] M. Comeche and J. Antonio, *Los modelos clásicos de Recuperación de información y su vigencia*: eprints.ucm.es, 2006.
- [57] A. Escárcega, "La lógica difusa en los sistemas automatizados de recuperación de información," *itescam.edu.mx*.
- [58] M. P. Molina, "BÚSQUEDA Y RECUPERACIÓN DE INFORMACIÓN " 2009.
- [59] G. Díaz and Raquel, "La evaluación en recuperación de la información," *recolecta.net*, 2006.
- [60] H. A. O. Erazo, "Meta Buscador Web Semántico Basado en una Taxonomía, Ontologías y Perfil de Usuario," 2010.
- [61] G. B. Carlos Cobos, Elvis Pérez, Martha Mendoza, "Modelo de un meta buscador web basado en taxonomías, ontologías y retroalimentación del usuario," 2009.
- [62] C. d'Amato, *Similarity-based Learning Methods for the Semantic Web*: PhD thesis, University of Bari, 2007.
- [63] L. Mazuel and N. Sabouret, "Semantic relatedness measure using object properties in an ontology," *The Semantic Web - ISWC 2008*, 2010.
- [64] R. Rada, *et al.*, "Development and application of a metric on semantic nets," *Systems, Man and Cybernetics, IEEE Transactions on*, 1989.
- [65] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *Arxiv preprint cmp-lg/9511007*, 1995.
- [66] J. Jiang and D. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," *Arxiv preprint cmp-lg/9709008*, 1997.
- [67] G. Hirst and D. St-Onge, *Lexical Chains as representation of context for the detection and correction malapropisms, chapter 13: WordNet: an electronic lexical database*, 1998.
- [68] T. Slimani, *et al.*, "A new similarity measure based on edge counting," *World Academy of Science, Engineering and Technology 23 2006*, 2004.
- [69] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," *ACL '94 Proceedings of the 32nd annual meeting on Association for Computational Linguistics* 1994.
- [70] J. Dorn and T. Naz, "Structuring meta-search research by design patterns," *International Computer Science and Technology Conference 2008*, 2008.
- [71] M. Manoj and E. J., "Information retrieval on Internet using meta-search engines: A review," *nopr.niscair.res.in*, 2008.
- [72] M. A. Niño, *et al.*, "UNICAUCA VIRTUAL: METAMODELOS DE UNIVERSIDAD VIRTUAL Y HERRAMIENTAS DE SOPORTE," *libra.niee.ufrgs.br*.
- [73] D. L. McGuinness and F. V. Harmelen, "OWL web ontology language overview," *W3C recommendation*, 2004.

- [74] M. A. Moreira and I. M. Greca, "Modelos mentales, modelos conceptuales y modelización," *Cuadernos catalanes de enseñanza de la física*, 1998.
- [75] U. d. Malaga, "Introducción al Modelado de Sistemas."
- [76] M. Baziz, *et al.*, "Evaluating a conceptual indexing method by utilizing WordNet," *Accessing Multilingual Information Repositories*, 2006.
- [77] A. Troussov, *et al.*, "Mining socio-semantic networks using spreading activation technique," *Proceedings of I-KNOW '08 and I-MEDIA '08 Graz, Austria, September 3-5, 2008*, 2008.
- [78] N. Karthikeyani and K. Thangavel, "Impact of normalization in distributed k-means clusterings," *Int. Journal of Soft Computing*, 2009.
- [79] A. Castillo, *et al.*, "Conceptualización del proceso de implementación de software: perspectivas ágil y disciplinada," *Ciencia e Ingeniería*, 2011.
- [80] LinkedDataTools. "Jena.NET Semantic Web Framework", [en línea] <http://www.linkeddatatools.com> [Consulta 09 de Octubre de 2010].
- [81] Bpellens. "OwlDotNetApi: Web Ontology Language API for the .NET platform", [en línea] <http://users.skynet.be/bpellens/OwlDotNetApi/owldotnetapi.html> [Consulta 09 de Octubre de 2010]
- [82] A. S. Foundation. "Lucene search engine library for .NET platform", [en línea] <http://incubator.apache.org/lucene.net/> [Consulta 29 de septiembre de 2011]
- [83] C. Gil, "Construcción de sitios web basados en el lenguaje de modelamiento UML y arquitectura multicapa," 2007.
- [84] CodePlex. "Facebook C# SDK: API to Integrate .Net Applications to Facebook", [en línea] <http://facebooksdk.codeplex.com/> [Consulta 30 de Octubre de 2010].
- [85] Microsoft. "ASP.NET AJAX Control Toolkit", [en línea] <http://www.asp.net/ajaxlibrary/AjaxControlToolkitSampleSite/> [Consulta 30 de septiembre de 2011].
- [86] j. Project, ""JavaScript Library", [en línea] <http://jquery.com/> [Consulta 30 de septiembre de 2011]."
- [87] C. Draganova, *Asynchronous javascript technology and xml (ajax)*: Citeseer, 2007.
- [88] N. Jenkins, *A Software Testing Primer*: nickjenkins.net, 2008.
- [89] J. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, Vol 76(5), 1971.
- [90] T. Joachims and F. Radlinski, "Search Engines that Learn from Implicit Feedback," *Computer*, vol. 40, pp. 34-40, 2007.
- [91] J. Landis and G. KochG, "The measurement of observer agreement for categorical data," 1977.